ISCR

Institute for Scientific Computing Research

Annual Report

Fiscal Year 2003

Lawrence Livermore National Laboratory, P.O. Box 808, L-561, Livermore, CA 94551

The University Relations Program (URP) encourages collaborative research between Lawrence Livermore National Laboratory (LLNL) and the University of California campuses. The Institute for Scientific Computing Research (ISCR) actively participates in such collaborative research, and this report details the Fiscal Year 2003 projects jointly served by URP and ISCR. For a full discussion of all URP projects in FY 2003, please request a copy of the URP FY 2003 Annual Report by contacting

Lawrence Livermore National Laboratory
Edie Rock, University Relations Program
P. O. Box 808, L-413
Livermore, CA 94551

UCRL-TR-202404

Work performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under Contract W-7405-ENG-48.

Contents

ISCR Institute for Scientific Computing Research

The Mission of the ISCR

The Institute for Scientific Computing Research (ISCR) at Lawrence Livermore National Laboratory (LLNL) identifies collaborators from the academic community for computer science and computational science research and development efforts at the Laboratory and hosts them in short- and long-term residential visits, with the aim of encouraging long-term academic research agendas that address Laboratory research priorities. Through such collaborations, ideas and software flow in both directions, and the Laboratory cultivates its future workforce. An extensively externally networked ISCR cost-effectively expands the level and scope of computational science expertise available to the Laboratory. As large-scale simulations on the parallel platforms of DOE's Advanced Simulation and Computing program (ASCI) and other terascale platforms become increasingly important to the overall mission of LLNL, the role of the ISCR expands in importance, accordingly.

The ISCR forms academic partnerships with the University of California (UC), with universities throughout the United States, and internationally. The primary consideration is to identify the highest quality, most topically opportune, and potentially most vigorous partnerships. The special relationship between LLNL and the University of California and the geographical proximity of UC's campuses tend to make collaborations with UC faculty and students particularly effective.

The ISCR strives to be the “eyes and ears” of the Laboratory in the computer and information sciences, keeping the Laboratory aware of and connected to important external advances. It also attempts to be “feet and hands” in carrying those advances into the Laboratory and incorporating them into practice. In addition to conducting research, the ISCR provides continuing educational opportunities to Laboratory personnel in the form of on-site workshops taught by experts on novel software or hardware technologies.

Through the workshops, visits, and internships it sponsors, ISCR also seeks to influence the research community external to the Laboratory to pursue Laboratory-related interests and to train the workforce that will be required by the Laboratory, because some of its needs are not otherwise well reflected in university curricula or the commercial information technology environment. Part of the performance of this function is interpreting to the external community appropriate (unclassified) aspects of the Laboratory's own contributions to the computer and information sciences, contributions that its unique mission and unique resources give it a unique opportunity and responsibility to make.

Institute for Scientific Computing Research
Fiscal Year 2003 Director's Report

Large-scale scientific computation and all of the disciplines that support it and help to validate it have been placed at the focus of Lawrence Livermore National Laboratory by the Advanced Simulation and Computing (ASCI) program of the National Nuclear Security Administration and the Scientific Discovery through Advanced Computing (SciDAC) initiative of the Office of Science of the Department of Energy (DOE). The Laboratory operates computers with among the highest peaks of performance in the world and has undertaken some of the largest and most compute-intensive simulations ever performed. Secretary of Energy Spencer Abraham announced in November 2003 that ultrascale simulation is one of the highest priorities in the DOE's facilities planning for the next two decades. However, computers at architectural extremes are notoriously difficult to use efficiently. Furthermore, each successful terascale simulation only points out the need for much better ways of interacting with the resulting avalanche of data.

Advances in scientific computing research have therefore never been more vital to the core missions of the Laboratory than at present. Computational science is evolving so rapidly along every one of its research fronts that to remain on the leading edge the Laboratory must engage researchers at many academic centers of excellence. In FY 2003, the Institute for Scientific Computing Research (ISCR) served as one of the Laboratory's main bridges to the academic community in the form of collaborative subcontracts, visiting faculty, student internships, workshops, and an active seminar series.

ISCR research participants are integrated into the Laboratory's Computing and Applied Research (CAR) Department, especially into its Center for Applied Scientific Computing (CASC). These organizations, in turn, address computational challenges arising throughout the rest of the Laboratory. Administratively, the ISCR flourishes under the Laboratory's University Relations Program (URP). Together with the other institutes of the URP, it navigates a course that allows the Laboratory to benefit from academic exchanges while preserving national security. While it is difficult to operate an academic-like research enterprise within the context of a national security laboratory, the results declare the challenges well met and worth the continued effort.

Fiscal year 2003 was the fourth full year under Acting Director David Keyes. Keyes, the Fu Foundation Professor of Applied Mathematic at Columbia University and an ISCR faculty participant since October 1997, dedicated one-third of his time to the technical program of the ISCR. Dr. James McGraw continued as the Deputy Director of the ISCR, while also serving in a special national role last year as the General Chair of the record-breaking SC 2003 conference. Linda Becker continued as the full-time Institute Administrator. Emma Horcabas and Leslie Bills logistically supported the large visitor and summer programs of the ISCR. They were assisted by Pam Mears, who joined the ISCR full-time.

The ISCR continues to have a small contingent of research staff members within its organization. Three ISCR staff—Prof. Nelson Max, Prof. Garry Rodrigue, and Prof. Rao Vemuri—hold joint appointments as professors at UC Davis and senior researchers at LLNL. In addition, the ISCR hosted eight post-doctoral staff: Alison Baker, David Buttler, Samson Cheung, Shawn Newsam, Dan Reynolds, Markus Schordan, Megan Thomas, and Qing Yi. Finally, the ISCR served as the host for ten students (listed in Table 1) who attend the University of California, Davis on a Student-Employee Graduate Fellowship. This fellowship enables students to work with LLNL researchers half-time while pursuing their PhDs.

The ISCR enables substantial interactions between academia and LLNL staff through consultants and participating guests. Consulting agreements are vehicles for permitting academics to interact with LLNL in a compensated fashion. Consultants can serve on special review committees, present classes on specific focus topics, and/or visit LLNL periodically for technical meetings. All consultants have a specific LLNL technical point of contact for overseeing the interactions. Table 2 lists the ISCR consultants for FY 2003. Participating Guests are researchers from academia or industry that often need intermittent access to LLNL staff (and resources), where funding is not the critical issue. This status permits an appropriate security clearance and the ability to quickly arrange for on-site visits with LLNL staff over a period of one month to two years. Table 3 lists ISCR's participating guests for FY 2003.

The pages of this report summarize the activities of the faculty members, post-doctoral researchers, students, and guests from industry and other laboratories who participated in LLNL's computational mission under the auspices of the ISCR during FY 2003. These activities fall under two main themes: sponsored research activities that stimulate interactions between academia and LLNL staff, and a diverse visitor program that enables both short- and long-term residential stays at LLNL.

ISCR oversees three different types of sponsored-research activities. The University Collaborative Research Program (UCRP), through the ISCR, funds typically six to nine research

Student	LLNL Advisor	Term at LLNL
Peer-timo Bremer	Dan Laney	6/02-6/06
Sam Brockington	Garry Rodrigue, Dave Hwang	10/01-8/06
Aaron Fisher	Garry Rodrigue	7/02-6/06
Benjamin Gregorski	Mark Duchaineau	6/01-6/05
Jeff Hagelberg	Paul Amala	9/03-9/04
Aaron Herrnstein	Michael Wickett	3/01-3/04
Ana Iontcheva	Panayot Vassilevski	9/00-8/03
Rob Rieben	Garry Rodrigue	10/00-9/04
Joshua Senecal	Mark Duchaineau	11/01-10/05
Yihao Zheng	Andy Wissink	7/02-8/03

Table 1: FY 2003 ISCR Student Employee Graduate Research Fellowships

projects each year at campuses of the University of California. These projects primarily support graduate students working on thesis research that focuses on a topic of interest to LLNL. The faculty principal investigators and students are expected to work closely with an LLNL collaborator. The ISCR also coordinates the funding of numerous research subcontracts to various academic institutions throughout the United States. These contracts are normally funded by programs at LLNL with the express intent of helping to address long-term Laboratory research interests. In many cases, this type of vehicle is used to fund sabbatical visits to LLNL for three to six months. With Laboratory Directed Research and Development (LDRD) funds, the ISCR also funds Exploratory Research in the Institutes (ERI). These research grants go to LLNL staff with the goal of developing ties to academia through co-funded research projects. Annual progress reports for UCRP-funded projects, subcontracts, and ERI projects can be found in the next three sections of this document.

In FY 2003 the ISCR continued its tradition of an extensive and diverse Visitor Program. This program includes sabbatical visitors, sponsored workshops, summer students, and three different seminar series featuring external speakers. The ASCI Institute

Consultant	Affiliation	LLNL Contact
Randolph Bank	UC San Diego	Charles Tong
Leo Breiman	UC Berkeley	Chandrika Kamath, Imola Fodor
Gene Golub	Stanford University	Edmond Chow
Anne Greenbaum	University of Washington	Peter Brown
Charles Hansen	University of Utah	Randy Frank
Heinz-Otto Kreiss	UC Los Angeles	David Brown
Thomas Manteuffel	University of Colorado, Boulder	Rob Falgout
Stephen McCormick	University of Colorado, Boulder	Rob Falgout
Linda Petzold	UC Santa Barbara	Peter Brown
Sandu Popescu	Bristol University	Patrice Turchi
Steve Schaffer	New Mexico Inst. of Mining and Technology	Rob Falgout, Jim Jones
Homer Walker	Worcester Polytechnic Institute	Carol Woodward

Table 2: FY 2003 ISCR Consultants

for Terascale Simulation Lecture Series was established to enrich the intellectual atmosphere of LLNL's large simulation community through the visits of leaders throughout the diverse areas of computation. In FY 2003 we hosted three speakers in this series. ISCR also co-hosted (with the Materials Research Institute) a special topical series on Quantum Computing that included 14 of the most prominent researchers in this emerging field. The general ISCR seminar series included an additional 43 talks covering a wide spectrum on research areas. Abstracts on all of these talks can be found in the Seminar Series section of this report.

Guest Name	Institution		
Marian Brezina	University of Colorado, Denver	Sally McKee	Cornell University
Alok Choudhary	Northwestern University	Michael Minion	University of North Carolina
Hans de Sterck	University of Colorado, Boulder	R. Frank Mueller	North Carolina State University
Branden E. Fitelson	UC Berkeley	Esmond Ng	Lawrence Berkeley Laboratory
Franz Franchetti	Technical University of Vienna	Beth Ong	LLNL (on leave)
Alejandro Garcia	San Jose State University	Peter Pacheco	University of San Francisco
Matthew R. Gibbons	U.S. Air Force Academy	Joseph Pasciak	Texas A&M University
Michael Griebel	University of Bonn	Joanne Perra	LLNL retired
Bernd Hamann	UC Davis	Christoph Pflaum	University of Erlangen
Alan Hindmarsh	LLNL (retired)	Elbridge Gerry Puckett	UC Davis
Ken Joy	UC Davis	Markus Pueschel	Carnegie Mellon University
Andrew Knyazev	University of Colorado, Denver	Ulrich Ruede	University of Erlangen
Johannes Kraus	University of Loeben	Paul Saylor	University of Illinois
Raytcho Lazarov	Texas A&M University	Martin Schultz	Cornell University
Lars Linsen	UC Davis	Claudio Silva	Oregon Graduate Institute
Oren Livne	Stanford University	Christoph W. Ueberhuber	Vienna University of Technology
Bertram Ludaescher	San Diego Supercomputer Center	Beata Winnicka	Argonne National Laboratory
Jennifer Mariani	UC Davis	Ludmil Zikatanov	Penn State University

Table 3: FY 2003 ISCR Participating Guests

Twenty-four faculty visitors were in residence for more than just a seminar visit—visits ranged from a week to a semester. Six of these faculty spent a portion of their sabbatical leave here. Altogether, the ISCR hosted 162 visits from 125 different visitors, an average of more than three visits per week. The vast majority of the visitors were from academia, with 17% from industry and 10% from other laboratories. Visitors from outside the United States made up 31% of the total.

During the summer, ISCR hosted 63 visiting students. The summer program exposes students to the stimulating and challenging work environment of a national laboratory. Successful candidates are hired as summer employees, assigned individual LLNL mentors, and given specific projects to which they will contribute. The nature of the project and contribution are determined so as to complement their background and skills. The topical coverage of the summer research program broadens each year as computation expands into new scientific areas and as computational tools become more powerful and diverse. Optimal algorithms, fluid turbulence, genomics, terascale visualization, and computer security are just a handful of topics from the lively summer hallway conversation at the ISCR! The summer program runs from mid-May to late September, with most participants spending 10–12 weeks on site. Project reports for most of the students can be found later in this report.

In June, with the advent of our large student summer program and sponsorship from the Defense Programs office of DOE Headquarters, we ramped up our fourth annual Internships in Terascale Simulation Technology Lecture Series. The tutors included David Brown, Erick Cantu-Paz, Terence Critchlow, Alex Garcia, Jeff Hittinger, Gary Kumfert, Carol Woodward, Kim Yates, CASC's Director Pete Eltgroth, and the ISCR Director. Though the lectures were intended for students, permanent CASC researchers attended an occasional subseries. A reporter from *Science* magazine attended the lecture by Kim Yates on LLNL's future BlueGene/L supercomputer and featured it in a subsequent article.

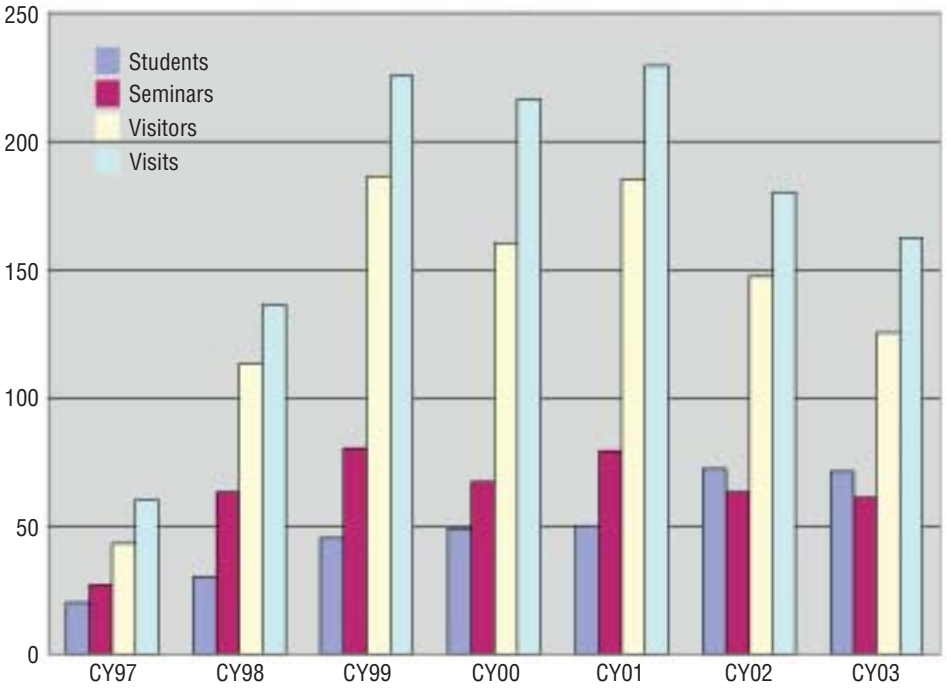


Fig. 1: ISCR Visitor Program CY1997–CY2003

Recognizing the growing number of computer scientists among the summer students, we added a second track of tutorial lectures during the summer months, called the ISCR Summer Lecture Series. The tutors included computer scientists Gary Kumfert and Kim Yates of CASC, who also

taught in the ITST Lecture Series, as well as Tony Bartoletti, Bill Cabot, Martin Casado, Bronis de Supinski, Paul Dubois, Tina Eliassi-Rad, Chandrika Kamath, Doug Lim, and Valerio Pascucci.

Computer Security specialist Terry Brugger also assembled a summer tutorial lecture series under the ISCR umbrella, under which were included the tutorials of Bartoletti, Cabot, and Casado mentioned above, plus several supplementary topics presented by Brugger, himself.

Figure 1 charts the numbers of visitors and seminars over the past seven years. The number of students in residence in FY'03 remained at its FY'02 high due to the expansion of the ISCR's responsibility in the larger CAR organization. The number of external seminars also remained about the same. (The seminar totals do not include the internally provided tutorial seminars, which doubled from 10 to 20.) The numbers of visits and visitors were slightly down, though the number of visitor-days was at the FY'02 level.

Most of the material of this annual report comes directly from the visitors and principal investigators of the projects being reported. We thank Dean Wheatcraft and Dale Sprouse for their editorial work and Dan Moore of the Technical Information Department of LLNL for his graphic artistry in producing an easily navigated and visually pleasing document.

We hope that you enjoy examining this report on the ISCR's diverse activities in FY 2003. For further information about the Institute, please contact us at the address below. Inquiries about how you might enhance the on-going FY 2004 program at the ISCR, or beyond, are welcome.

David E. Keyes

Institute for Scientific Computing Research
Lawrence Livermore National Laboratory
P. O. Box 808, L-561
Livermore, CA 94551
http://www.llnl.gov/iscr

David E. Keyes
Acting Director
http://www.llnl.gov/iscr/people/keyes
dekeyes@llnl.gov
925-422-1325

James R. McGraw
Deputy Director
http://www.llnl.gov/iscr/people/mcgraw
jmcgraw@llnl.gov
925-422-0541

Linda Becker
Administrator
http://www.llnl.gov/iscr/people/becker
becker22@llnl.gov
925-423-0421

Institute for Scientific Computing Research

Institute for Terascale Simulation Lecture Series

(in reverse cronological order)

Institute for Terascale Simulation Lectures

The ASCI Institute for Terascale Simulation Lecture Series was established to enrich the intellectual atmosphere of LLNL's large simulation community through the visits of leaders throughout the diverse areas of computation that undergird simulation. Simulation has become a crucial third mode of scientific investigation and engineering design, along with theory and experiment, and has become especially important for fundamental scientific progress and technical decision support to the U.S. Department of Energy under the Advanced Simulation and Computing Program (ASCI). ITS lectures are designed to appeal to a broad technical audience and are open to all Laboratory staff.

September 11, 2003

WebBase: Building a Web Warehouse

Hector Garcia-Molina
Stanford University

Email: hector@cs.stanford.edu

Abstract

The World Wide Web, or simply the Web, is rapidly becoming the world's collective information store, containing everything from news to entertainment, personal communications, and product descriptions. This world information store is distributed across millions of computers, but it is often important to gather significant parts of it at a single site. This is one reason is to build content indices, such as Google. Another reason is to mine the cached Web, looking for trends or data correlations. A third reason for gathering a Web copy is to create a historical record for Web sites that are ephemeral or changing. In this talk, I will discuss how to build a repository of Web pages, describing some of the technical challenges faced. I will illustrate with some of the work we have been doing in our group at Stanford.

Speaker's web page: http://www-db.stanford.edu/people/hector.html

June 6, 2003

Modeling Climate and Future Climate Change

Warren Washington
National Center for Atmospheric Research

Email: wmw@ucar.edu

Abstract

Everyone knows that climate has always changed. So what is unique about what has happened over the last century or so? Many climate research scientists now believe that humans are changing Earth's system and that global warming is taking place. Other scientists are skeptical of this view and think the observed changes result from natural climate variability. A review of recently observed climate change will be presented and compared with climate model simulations of the recent past climate and what is expected in the 21st century. A discussion of the uncertainties will be included along with an analysis of policy options. The future for terascale computing of climate will also be discussed.

Speaker's web page: http://www.cgd.ucar.edu/ccr/warren/

April 14, 2003

A New Kind of Science

Stephen Wolfram
Wolfram Research, Inc.

Email: contact@wolframscience.com

ISCR Institute for Scientific Computing Research

Abstract

Starting from a few computer experiments, Stephen Wolfram has spent more than 20 years developing a new approach to science, described for the first time in his book *A New Kind of Science*. Basic to his approach is the idea of studying not traditional mathematical equations but rules of the kind embodied in the simplest computer programs. A key discovery is that such rules can lead to behavior that shows immense complexity and mirrors many features seen in nature. Wolfram has built on this to tackle a remarkable array of fundamental problems in science, from the origins of apparent randomness in physical systems, to the development of complexity in biology, the ultimate scope and limitations of mathematics, the possibility of a truly fundamental theory of physics, the interplay between free will and determinism, and the character of intelligence in the universe. Released on May 14, 2002, Wolfram's book became an instant bestseller and is now showing many signs of initiating a major paradigm shift in science. Wolfram's presentation will cover some of the key ideas and discoveries in his book, outlining their implications, and discussing their personal and historical context. An extended question and answer period will be included.

Speaker's web page: http://www.wolframscience.com/events/

Institute for Scientific Computing Research

Quantum Computation and Information Lecture Series

(in reverse cronological order)

Quantum Computation and Information Seminars

The Quantum Computation and Information Seminar Series was established to provide a perspective and highlights of areas of quantum computation and quantum information useful to LLNL. Held over the course of a year, it featured visiting experts addressing many aspects of quantum information science. This series was co-hosted by the ISCR and the Materials Research Institute at LLNL.

(*Anthony Leggett and Charles Bennett were part of the QCI Seminar Series but hosted primarily by the LLNL Director through the Director's Distinguished Lecture Series.)

May 5, 2003

Quantum Information Science and Technology—Defense Advanced Research Projects Agency's (DARPA's) Vision

Stu Wolf
University of Virginia

ISCR Institute for Scientific Computing Research

Abstract

The potential advantages of using quantum mechanical effects in computing and communication have been known for some time. However, up to now, little focused research has addressed the critical issues that would allow the Department of Defense (DoD) communications and/or computational applications to benefit from that potential. QUIST is a comprehensive program designed to address these issues by developing an understanding of the science and theory of quantum information and processing through the demonstration of critical implementations. Specifically, this program will involve the formulation of new algorithms and protocols for ultrasecure (zero probability of intercept) communications. In addition, QUIST will develop and validate new algorithms and methods for solving "quantum complex" problems of importance to DoD—for example, the problem of graph isomorphism, which may have important implications for cryptography and error correction. This program will also explore the limits of quantum computation for dramatically improved speedups of classical computations. Concurrent with these advances, QUIST will develop the component technology for secure quantum communication, including the development of robust megahertz-rate single photon sources and detectors, single quantum bit (qubit) gates, two qubit-controlled NOT gates, and quantum memory. This technology will allow the demonstration of quantum teleportation and ultrasecure communication over long distances (100 km). The development of this technology will provide some of the building blocks for quantum computation, which will require many more quantum gates and larger memory.

April 10, 2003

Algorithms for Quantum Computers

David Meyer
Universiity of California, San Diego

Abstract

Interest in quantum computing exploded after Shor discovered a quantum algorithm to factor numbers exponentially faster than the best classical algorithm known. Despite the importance of this result for the (in)security of classical cryptosystems, other problems with superior quantum solutions must exist to justify the immense commitment of resources that will be necessary to build a quantum computer. In this talk, I will explain Grover's quantum search algorithm and related algorithms with potential applications to problems of interest at Lawrence Livermore National Laboratory, specifically, solving NP problems, simulating classical physical systems, and image processing.

April 7, 2003

Decoherence and the Problem of Implementing Quantum Computation

Roland Omnés
University Paris-Sud

ISCR Institute for Scientific Computing Research

Abstract

Decoherence is the most troublesome problem in the field of quantum computing. It turns out that the close relation between irreversible phenomena and decoherence yields the current most general theory of decoherence. This relation solves a problem about the existence of a diagonalization basis (pointer basis) of importance for the understanding of decoherence. I will try to make clear the nature of the decoherence, its history, the main problems it raises and their present status, as well as its possible consequences for the prospects of quantum computing.

March 27, 2003

Quantum Information and Decoherence Studies in Cavity QED Experiments

Serge Haroche

Ecole Normale Supérieure and Collège de France

Abstract

In cavity quantum electrodynamics (QED) experiments, Rydberg atoms and single microwave photons are used as qubits. Quantum gates based on resonant and dispersive atom-field effects have been realized, which implement various kinds of conditional dynamics between these qubits. Three-particle entanglement has been achieved by combining two-qubit operations. A scheme implementing Grover's search algorithms has been proposed. Generalizing these studies to larger systems involving more atoms and photons is challenging. Decoherence produced by field relaxation in the cavity is the most fundamental and difficult effect to control. After a general presentation about our apparatus, we will review these experiments and discuss their possible development, which implies in the short-term improvements in our cavity design and quantitative decoherence studies in the upgraded apparatus. In longer-term projects, we plan to extend our experiments to two cavities and study nonlocal quantum effects involving mesoscopic field superpositions.

February 24, 2003

Cryptography in the Quantum Computing Era

Claude Crépeau
McGill University

ISCR Institute for Scientific Computing Research

Abstract

Introductions to classical cryptographic tools and quantum computing will first be presented. The basic notion of quantum key distribution will then be discussed. Information on theoretical notions of cryptography over quantum states such as encryption and authentication will be covered. Computational analogues will also be presented: quantum public-key cryptography, public-key authentication, and impossibility of quantum digital signatures. Some applications to quantum error-correcting codes will be presented.

February 11, 2003

Quantum Information

Charles H. Bennett
IBM T. J. Watson Research Center

Abstract

Information and computation theory have recently been extended to include the transmission and processing of intact quantum states and feats such as quantum cryptography, quantum teleportation, and fast quantum computation. Although progress toward a practical quantum computer is slow, the theory of quantum information processing has developed to the point where it can be viewed as the most natural and complete formulation of the notions of information and computation. Quantum information processing extends classical information and computation theory in much the way that complex numbers extend real numbers. In this lecture, we review the capacities of quantum channels and the use of auxiliary resources, such as shared entanglement, in the transmission of classical and quantum information.

January 27, 2003

So You've Built a Quantum Computer—Now What Are You Going to Do with It?

Seth Lloyd
Massachusetts Institute of Technology

ISCR Institute for Scientific Computing Research

Abstract

This talk details recent advances in quantum computation and quantum communication. A variety of technologies now exist for constructing simple quantum computers and quantum communication systems. Examples include NMR, quantum optics, and superconducting systems. But once a quantum information processing device has been constructed, what do users do with it? This talk discusses applications of quantum information processing, including quantum algorithms, quantum analog computing, quantum games, quantum internet protocols, and quantum weirdness.

January 21, 2003

Quantum Algorithms: From Quantum Interference to Quantum Annealing

Artur Ekert
University of Cambridge

Abstract

The theory of classical universal computation was laid down in 1936, was implemented within a decade, became commercial within another decade, and dominated the world's economy half a century later. Quantum information technology is a fundamentally new way of harnessing nature. It is too early to say how important this technology will eventually be, but we can reasonably speculate about its effect on computation. Quantum computers use the quantum interference of different computational paths to enhance correct outcomes and suppress erroneous outcomes of computations. A common pattern underpinning quantum algorithms can be identified when quantum computation is viewed as multiparticle interference. I will use this approach to review some of the existing quantum algorithms and to outline new concepts and architectures for implementing quantum computation.

January 10, 2003

Manipulating Quantum Information with Semiconductor Spintronics

David D. Awchalom
University of California,
Santa Barbara

ISCR Institute for Scientific Computing Research

Abstract

There is a growing interest in the use of electronic and nuclear spins in semiconductor nanostructures as a medium for the manipulation and storage of classical and quantum information. Spin-based electronics offer remarkable opportunities for exploiting the robustness of quantum spin states by combining standard electronics with spin-dependent effects that arise from the interactions between electrons, nuclei, and magnetic fields. In this lecture, we provide an overview of recent developments in coherent electronic spin dynamics in semiconductors and quantum structures, including a discussion of temporally and spatially resolved magnetooptical measurements that reveal an interesting interplay between electronic and nuclear spins. These experiments explore the electronic, photonic, and magnetic control of electron and nuclear spins in a variety of nanostructures and focus on investigating the underlying physics for quantum information processing in the solid state.

December 19, 2002

Quantum Nonlocality

Sandu Popescu
University of Bristol

Abstract

One of the most exotic aspects of the behavior of microscopic particles is the so-called quantum nonlocality, or entanglement. Microscopic particles—or other larger quantum systems—that have interacted in the past and then moved far from each other remain, in a certain sense, connected and can instantaneously "communicate" with each other (in apparent but not real contradiction with Einstein's relativity). Nonlocality is now considered to be the basic ingredient in the newly developed area of quantum information and computation. In my talk, I will explain at a very accessible level the basic idea of quantum nonlocality, and I will discuss some of its applications.

December 12, 2002

Quantum Computation with Trapped Atomic Ions

David J. Wineland

National Institute of Standards and Technology

ISCR Institute for Scientific Computing Research

Abstract

In spite of considerable interest in the possibility of making a quantum computer, it is generally agreed that building one capable of useful factoring or database searching will be extremely difficult. Atomic physics experiments can satisfy many of the requirements for a quantum computer, including a path to realize a large-scale device; however, difficult technical problems must be overcome. Some of these problems, in the context of trapped atomic ions, will be discussed. In the meantime, the ideas of quantum information processing have clarified our thinking about simpler tasks, some of which can now be implemented. For example, simple quantum processing can now increase the signal-to-noise ratio in spectroscopy and may broaden the choices of atoms that can be used for atomic clocks.

December 9, 2002

Quantum Information Science—The Promise, the Problems, and the Plumbing

Jeff H. Kimble
California Institute of Technology

Abstract

In recent years, a remarkable set of advances has occurred at the interface of information science and quantum mechanics, spanning from communication to computation. Quite unexpectedly, certain tasks that are otherwise impossible in the classical world become possible in the quantum realm. This lecture will give an overview of these "quantum miracles," as well as of the discovery of quantum error correction and fault tolerance that enable reliable computation with imperfect components. A survey of the diverse physical systems that are being explored for the implementation of quantum logic will be presented, with an emphasis on the research at the California Institute of Technology of strong coupling of single atoms and photons, including the realization of quantum teleportation.

November 4, 2002

The Nature and Location of Quantum Information

Robert B. Griffiths
Carnegie Mellon University

ISCR Institute for Scientific Computing Research

Abstract

Extending the classical information theory of Shannon into the quantum domain runs into the difficulty that textbook quantum theory has no consistent scheme for assigning probabilities to microscopic objects in the absence of measurements. I will discuss how to get around this problem and construct a theory that provides precise answers to questions such as: How many bits of information can be carried by a single qubit? Where is this information located during processes of dense coding and teleportation? Can quantum information travel backward in time? In each case, classical analogies can help us understand the quantum results.

October 22, 2002

Bell's Theorem, Entanglement, Teleportation, Quantum Computing and All That

Anthony Leggett
University of Illinois, Champaign–Urbana

Abstract

One of the most surprising aspects of quantum mechanics is that under certain circumstances, it does not allow individual physical systems, even when isolated, to possess properties of their own. In the three decades since John Bell first clearly appreciated quantum mechanics revolutionary significance in 1964, this feature has been tested experimentally and spectacularly confirmed, in the opinion of most. More recent discoveries show that it facilitates certain operations that are classically impossible, such as teleportation, secure-in-principle cryptography, and quantum computing (at least in principle). This talk gives a basic introduction to the subject and looks at recent advances that suggest that relatively macroscopic systems (Josephson devices) may be viable candidates for the elements—called qubits—of a quantum computer.

Institute for Scientific Computing Research

Seminar Series Abstracts

(in reverse cronological order)

ISCR Seminar Series

The ISCR hosts on-site seminars by visiting experts at the request of Laboratory researchers. Approximately one seminar is held per week throughout the year.

August 29, 2003

Numerical Conservation Properties of Least-Squares Finite Element Methods

Hans de Sterck
University of Colorado at Boulder

Email: desterck@newton.colorado.edu

Abstract

This talk presents results from our ongoing study of least-squares finite element methods (LSFEMs) for nonlinear hyperbolic conservation laws. The conservation law is reformulated by introducing the flux vector, or the associated flux potential, explicitly as additional dependent variables. The standard least-squares finite element procedure is then applied to the reformulated equations using H(div)-conforming finite element spaces and a Gauss–Newton nonlinear solution technique with grid continuation. Numerical results are presented for the one-dimensional Burgers equation on adaptively refined space–time domains, indicating that the H(div)-conforming finite element methods converge to the entropy-weak solution of the conservation law. The H(div)-conforming LSFEMs do not satisfy a discrete exact conservation property in the sense of Lax and Wendroff. However, weak conservation theorems that are analogous to the Lax–Wendroff theorem for conservative finite difference methods are proved for the H(div)-conforming LSFEMs. These results illustrate that discrete exact conservation in the sense of Lax and Wendroff is not a necessary condition for numerical conservation but can be replaced by minimization in a suitable continuous norm. This is joint research with Luke Olson, Tom Manteuffel, and Steve McCormick.

Speaker's web page: http://amath.colorado.edu/faculty/desterck

Institution web page: http://www.colorado.edu/

August 14, 2003

A Framework for Practical Applications Performance Modeling

Michael McCracken
University of California, San Diego

Email: mccrack@cs.ucsd.edu

Abstract

This talk describes technical advances that have enabled a framework for performance modeling to become practical for analyzing the performance of HPC applications; also, studies of some applications are exhibited to indicate the kinds of insights the framework can provide. In previous work, we presented a framework for performance modeling and prediction that is faster than cycle-accurate simulation and more informative than simple benchmarking and that was shown useful for performance investigations via proof-of-principle studies applied to short-running kernels. Here, we provide an update on the subsequent investigations we have carried out to advance the framework. The methods are now extensible to longer-running applications. The dominant cost in the framework has always been the time required to gather "application signatures" via instrumented application traces. We describe how we reduced tracing time while preserving the accuracy of resulting performance models. The framework is used to model the performance of three applications, POP (Parallel Ocean Program), NLOM (Navy Layered Ocean Model), and Cobalt60 run with exemplary data inputs at various processor counts on several HPC platforms (with Power3, Power4, Alpha, and Cray X1 processors). We show how the models can be used to develop sensitivity profiles of these applications to quantify how their performance would increase from improvements in the underlying architecture. This research was done in collaboration with Allan Snavely.

Speaker's web page: http://www.cs.ucsd.edu/~mmccrack/

Research web page: http://www.sdsc.edu/PMaC/

Institution web page: http://www.ucsd.edu/

August 11, 2003

Query Processing in Sensor Networks

Sam Madden
University of California, Berkeley

Email: madden@cs.berkeley.edu

Abstract

Many of the emerging applications for sensor networks are focused on data collection and monitoring in remote environments. Unfortunately, existing tools for building such applications require users of these networks, who often are not trained computer scientists, to write low-level, embedded C code. Such deployments frequently become mired in the difficulties of coding power-management, routing, and storage features in these volatile distributed environments.

In this talk, I will discuss how many of these difficulties can be overcome by providing users with a simple declarative interface where short, SQL-like queries are pushed into the network. Such queries concisely express a user's data needs, freeing him or her from the details of implementation and execution. In addition to dramatically simplifying the task of sensor-network programming, this approach enables the system to transparently optimize in-network query execution to minimize overall power consumption in ways that even sophisticated programmers may miss.

I will summarize the query processing features of TinyDB, a query processor for sensor networks we have developed at Berkeley, focusing on a framework for executing and optimizing aggregation queries. I will discuss current deployments that are under way at Berkeley, along with new features that are being incorporated to accommodate these deployments. I will include a brief demonstration of the system.

Speaker's web page: http://www.cs.berkeley.edu/~madden/berkeley.htm

Institution web page: http://www.berkeley.edu/

August 8, 2003

Automated Comparative Profiling of Parallel Cluster Applications

Karen Karavanic
Portland State University

Email: karavan@cs.pdx.edu

Abstract

The PPerfDB research project is developing scalable methods for diagnosing the performance of large-scale parallel applications using data from more than one program execution. PPerfXchange, a tool component under development, allows geographically dispersed data stores to be included as sources of data for PPerfDB. PPerfXchange models each site's performance data as XML documents, based upon a global schema, allowing client applications to retrieve performance data from remote sites using XQuery.

To gain insights for our continued work developing PPerfDB, we recently completed a comparative performance study of two MPI implementations on a dual-processor Linux cluster. I will present the results from this study and discuss our conclusions. I will also present our ongoing research into the issues related to storing and sharing performance data as part of an automated performance diagnosis environment.

Speaker's web page: http://www.cs.pdx.edu/~karavan/research.html

Institution web page: http://www.pdx.edu/

August 7, 2003

DNS of Particle-Laden Turbulent Flows

Said Elghobashi
University of California, Irvine

Email: selghoba@uci.edu

Abstract

Turbulent flows laden with particles are encountered in nature and numerous engineering applications. When the volume fraction of the dispersed phase is large enough, the turbulence properties (e.g., kinetic energy and dissipation spectra, strain rates) of the carrier flow undergo modifications, which, in turn, affect the dispersion characteristics of the particles (or bubbles). The objective of this seminar is to discuss the results of direct numerical simulations of two flows—particle-laden isotropic turbulence and particle-laden turbulent homogeneous shear flows—with emphasis on the physical mechanisms responsible for turbulence modifications.

Speaker's web page: http://kolmog.eng.uci.edu/

Institution web page: http://www.uci.edu

July 25, 2003

Sparse-Grid and Multigrid Methods for Multidimensional Problems

Christoph Reisinger
University of Heidelberg

Email: christoph.reisinger@iwr.uni-heidelberg.de

ISCR Institute for Scientific Computing Research

Abstract

Recently, there has been an increasing focus on problems where the dimensionality exceeds that which can be accommodated by standard tensor-product grid representations. Prominent examples include quantum mechanics (Schroedinger equation) and option pricing (Black–Scholes equation).

To deal with the "curse of dimensionality" encountered on, for example, Cartesian grids, we represent the solution on so-called sparse grids. In the combination technique, the solution is then extrapolated from small anisotropic grids with log-linear complexity. We can show that the convergence deteriorates only by logarithmic factors as well.

For a scalable method in a parallel setting, this demands from the linear (multigrid) solver not only optimal (linear) complexity that is robust with respect to the underlying equations, but also the anisotropy of the grids. Numerical results for block relaxation methods are shown for parabolic equations and free-boundary-value problems.

Institution web page: http://www.uni-heidelberg.de/index_e.html

July 18, 2003

Compiler Management of Global and Dynamic Data Reuse

Chen Ding
University of Rochester

Email: cding@cs.rochester.edu

Abstract

Reusing data in cache is critical to achieving high performance on modern machines. Many programs reuse a large amount of data in changing patterns across the whole program. Examples include an astronomical simulation of a changing galaxy in time steps, a compiler pulling together a set of functions in multiple passes, and a game-playing program calculating a sequence of moves through many searches. The question our work addresses is how well one can analyze and organize data reuses across large, dynamic program phases.

This presentation will describe several studies we conducted in the past seven years. First, I will describe a two-step strategy for improving global and dynamic cache reuse. The first step fuses computation on the same data to enable caching of repeated accesses. The second step groups data used by the same computation to effect contiguous access to memory. I will show how much we can automate this strategy across the whole program, over the entire data, and throughout program execution. Next, I will present recent results in predicting reuse distance pattern and demonstrate its use with a 3D web applet. For a wide range of benchmark programs, the applet predicts the cache miss rate and its changes for all cache sizes and all program inputs. Throughout the talk, I will also review classical studies that have laid the foundation for this and other work in memory hierarchy management.

Speaker's web page: http://www.cs.rochester.edu/~cding/

Research web page: http://www.cs.rochester.edu/~cding/Research

Institution web page: http://www.rochester.edu/

July 17, 2003

A Domain Decomposition Solver for a Parallel Adaptive Meshing Paradigm

Randolph Bank
University of California, San Diego

Email: reb@sdna1.ucsd.edu

Abstract

We describe a domain decomposition algorithm for use in the parallel adaptive meshing paradigm of Bank and Holst. Our algorithm has low communication, makes extensive use of existing sequential solvers, and exploits in several important ways data generated as part of the adaptive meshing paradigm. Numerical examples illustrate the effectiveness of the procedure. This work was done in collaboration with Shaoying Lu.

Speaker's web page: http://www.scicomp.ucsd.edu/~reb/

Research web page: http://www.scicomp.ucsd.edu/~reb/

Institution web page: http://www.scicomp.ucsd.edu/

July 11, 2003

Nonlinear Optimization Framework for Image-Based Modeling on Programmable Graphics Hardware

Radek Grzeszczuk
Intel Corporation

Email: Radek.Grzeszczuk@intel.com

Abstract

Graphics hardware is undergoing a change from fixed-function pipelines to more programmable organizations that resemble general purpose stream processors. In this paper, we show that certain general algorithms, not normally associated with computer graphics, can be mapped to such designs. Specifically, we cast nonlinear optimization as a data-streaming process that is well matched to modern graphics processors. Our framework is particularly well suited for solving image-based modeling problems since it can be used to represent a large and diverse class of these problems using a common formulation. We successfully apply this approach to two distinct image-based modeling problems: light field mapping approximation and fitting the Lafortune model to spatial bidirectional reflectance distribution functions. Comparing the performance of the graphics hardware implementation to a CPU implementation, we show a more than 5-fold improvement.

Speaker's web page: http://www.intel.com/research/people/bios/grzeszczuk_r.htm

Research web page: http://www.intel.com/research/

Institution web page: http://www.intel.com

June 19, 2003

PC-Based Visualization: Hardware and Software Solutions

Charles Hansen
University of Utah

Email: hansen@cs.utah.edu

ISCR Institute for Scientific Computing Research

Abstract

Clusters of PCs provide the opportunity for high-performance visualization methods of large-scale simulation data at lower cost than currently available. This talk will focus on three projects at Utah that leverage a PC cluster or PC graphics programmability applicable to clusters.

Real-time ray-tracing for isosurfacing has proven to be the most interactive method for large-scale scientific data. However, the limitation of requiring a tightly coupled DSM computer with a large number of processors has led to a bottleneck in the usability of such methods. We have developed a mechanism for allowing PC clusters to perform the same tasks as a tightly coupled Origin system. In particular, we are able to render the Lawrence Livermore National Laboratory Richmeyer–Meshkov instability dataset at multiple frames per second. Illumination in volume rendering has been limited because of the standardization on the ancient Phong shading model. We have developed a translucency model for volumetric rendering. Recently, we have extended this work to include an evaluation of a Gaussian approximation to the rendering equation that we call the Gaussian transfer function method. This is similar to preintegrated volume rendering for multidimensional transfer functions. The ability to interactively segment volumes for feature extraction and region-of-interest analysis is possible with the programmability of the modern graphics cards (GPU). We have developed a method for interactive level-set computation for segmentation and coupled this with real-time volume rendering. The contribution is the applicability of sparse methods to a coupled GPU and CPU implementation.

Speaker's web page: http://www.cs.utah.edu/~hansen/

Institution web page: http://www.utah.edu/

June 11, 2003

The Mathematica Platform

Andrew de Laix
Wolfram Research, Inc.

Email: delaix@wolfram.com

Abstract

Mathematica is a comprehensive platform for software development that combines a large library of algorithms, an advanced programming language, and an interactive notebook interface into a single package. This talk will offer a conceptual overview of Mathematica, focusing on its key components: the algorithm knowledge base, the symbolic programming language, and the document-centered interface. The algorithm knowledge base contains over 4000 functions forming a wide-ranging toolbox for building software. The symbolic programming language offers high-level constructs for manipulating not only numbers and arrays, but also complex tree structures, allowing one to write compact and flexible code. Mathematica's notebook interface enables one to combine text, mathematical typesetting, interactive elements, and code into a single document. Software, documentation, graphics, and interactive examples can be combined into a notebook for a unified presentation. Together, these elements constitute a flexible platform for developing a wide range of software solutions.

Institution web page: http://www.wolfram.com/

May 15, 2003

Scientific Data Management: A Case for Using High-level Information for I/O Optimizations

Alok Choudhary
Northwestern University

Email: choudhar@ece.northwestern.edu

Abstract

Management, storage, and efficient access to and analysis of huge amounts of data that are likely to be generated and/or used in various phases of large-scale scientific experiments and simulations are challenging tasks. Current data management and analysis techniques do not measure up to the challenges posed by such requirements in term of performance, scalability, ease of use, and interfaces.

In this talk, we will describe the research and development at Northwestern University to address the above problems. We will present optimizations techniques that use high-level access pattern information to enable optimizations in parallel file accesses. In particular, we describe techniques that can reduce or eliminate locking by explicitly managing I/O concurrency, which is an important step in achieving true scalability. We present performance results by incorporating these techniques in MPI-IO layer. In the second part, we will present the architecture and implementation of a metadata management system that allows the user to store, analyze, and use access patterns, relationships among data sets, data analysis, and I/O optimizations for scientific applications. We will describe the use of automatic I/O optimization techniques that can be incorporated into applications in a seamless fashion. Finally, we present recent results on data mining techniques that we are developing for online data analysis for scientific applications.

Speaker's web page: http://www.ece.northwestern.edu/~choudhar

Institution web page: http://www.northwestern.edu/

May 13, 2003

Automated Reconstruction of 3D City Models By Merging Ground-Based and Airborne Views

Avideh Zakhor
&
Christian Früh
University of California, Berkeley

Email: avz@eecs.berkeley.edu, frueh@eecs.berkeley.edu

Abstract

Three-dimensional (3D) models of urban environments are useful in a variety of applications such as urban planning, computer games, special effects for the movie industry, training and simulation for urban terrorism scenarios, and virtual heritage conservation. Currently, acquisition of 3D city models is difficult and time consuming; hence, commercially available models typically take months to create, usually require significant manual intervention, and lack the level of detail necessary for many applications.

We present a fast approach to completely automated reconstruction of textured 3D city models with both high details at ground level and complete coverage for bird's-eye view. While driving under normal traffic conditions on public roads, a vehicle equipped with two 2D laser scanners and a digital camera can acquire close-range laser scans and images of the facades at the ground level. Furthermore, a far-range digital surface map (DSM) containing complementary roof and terrain shape is created from airborne laser scans, then triangulated, and finally texture mapped with aerial imagery. While an initial estimate for the acquisition vehicle's motion is derived from scan-to-scan matching, the final pose with respect to the DSM is determined by using Monte Carlo localization, hence globally registering the facade data with the DSM. We have developed a framework of data processing algorithms that copes with imperfections inherent in city laser scans such as foreground occlusions and reflections from glass surfaces and is capable of reconstructing both geometry and texture of the facades. The resulting facade models and the airborne mesh from the DSM are merged to a single consistent model by removing redundant parts and filling gaps. The developed algorithms are evaluated on a data set acquired in downtown Berkeley, California, and we present the resulting textured 3D city model.

Avideh Zakhor web page: http://www-video.eecs.berkeley.edu/~avz

Christian Früh web page: http://www-video.eecs.berkeley.edu/~frueh/

Research web page: http://www-video.eecs.berkeley.edu

Institution web page: http://www.berkeley.edu/

April 30, 2003

Adaptive Mesh Refinement for Miscible Displacement

John Trangenstein
Duke University

Email: johnt@math.duke.edu

Abstract

Multicomponent flow in porous media with random heterogeneity often involves concentration fronts that are difficult to resolve with finite difference methods. In order to concentrate the computational effort in modeling such problems, we will describe a dynamically adaptive mesh refinement algorithm.

The flow equations are of two types: convection-dominated mass conservation equations and an elliptic pressure equation representing a combination of Darcy's law and incompressibility. The mass conservation equations are integrated by formally second-order accurate donor cell upwind finite difference schemes. One such scheme, a modification of the corner-transport upwind scheme given an incompressible velocity field, is being developed with help from Jeffrey Hittinger. The pressure equation is discretized by a hybrid mixed finite element method. The permeability field is generated by a Markov random field.

Application of these numerical methods to hierarchical adaptive grids requires the application of several discrete numerical conservation laws for communication between scales. However, iterative solution of the linear system for the pressure field on the adaptive grid is significantly more complicated.

We will describe an iterative strategy that uses a conjugate gradient outer iteration, with multiplicative domain decomposition as its preconditioner. Multiplicative domain decomposition (which is similar to multigrid) is applied between levels of refinement. Within a level of refinement, we use a block Jacobi iteration equivalent to solve $2 \times 2 \times 2$ elliptic boundary value problems for the multigrid smoother. The 3D multigrid algorithm was completed and debugged with help from Robert Falgout and Panayot Vassilevski.

Numerical results for viscous fingering in 2D and 3D will be presented. The code has been parallelized, with significant help and code from Andrew Wissink and David Hysom. These results are being compared to the Todd–Longstaff mixing model to develop models of the error in its prediction of effective mixing. This work is being conducted in collaboration with Mike Christie of Heriot-Watt University.

Speaker's web page: http://www2.math.duke.edu/faculty/johnt

Institution web page: http://www.duke.edu/

April 14, 2003

What We Have Here Is a Failure to Communicate: Application Integration in Emergency Services

Bruce Rankin
City of Edmonton, Alberta, Canada

Email: Bruce.Rankin@edmonton.ca

Abstract

This presentation of the current state-of-the-art information systems for emergency first-response services gives insight into what works well in the real world and where there is still work to be done. The first surprising revelation is that no single information management system is used to respond to emergencies, Rather, a collection of systems is used by fire and emergency management services to get the needed information to the right people. A brief operations overview will describe the functionality required by these systems and provide a backdrop to a technical discussion outlining their capabilities. The technical discussion will focus on two examples used throughout the presentation: what happens during a 911 call and what happens during an extended emergency situation. These examples will highlight how information systems support logistical, administrative, operational, and strategic purposes. Integration issues will also be discussed, and I will conclude with goals for future computing environments.

Institution web page: http://www.edmonton.ca/

April 7, 2003

Supporting Soft Real-Time Processing in Best-Effort Systems

Scott Banachowski
University of California, Santa Cruz

Email: sbanacho@cse.ucsc.edu

Abstract

General-purpose computing platforms are commonly used to execute time-constrained applications—for example, running a multimedia player on a desktop system. These systems use best-effort scheduling algorithms that make no guarantees about the timing of resource allocation. The best-effort model is attractive for both developers and users because it doesn't require any interface for negotiating resource allocations. For the same reason, best-effort systems lack support for tasks with deadlines, limiting their usefulness to situations where the available resources far exceed the demands of applications.

This talk focuses on supporting soft real-time processing within the best-effort model. We present online techniques for inferring real-time behavior from tasks and improving responsiveness for those with periodic deadlines. The advantage of this approach is that a best-effort system can automatically provide soft real-time support without intervention from users or prior knowledge of tasks' timing constraints. We have created two CPU schedulers based on this technique that combine desirable aspects of both best-effort and soft real-time scheduling. The first scheduler provides accurate estimation of deadlines, and the second scheduler reduces scheduling latency while preserving the time-share notion of fairness.

Speaker's web page: http://www.soe.ucsc.edu/~sbanacho/

Institution web page: http://www.ucsc.edu/

April 4, 2003

An HLLC-Type Approximate Riemann Solver for Ideal Magnetohydrodynamics

Katharine Gurski
National Institute of Standards and Technology

Email: kgurski@nist.gov

Abstract

In the numerical simulation of magnetohydrodynamics (MHD), a balance is necessary between capturing the key features of the flow, limiting computational expense, and the robustness of the numerical method. This talk will present a method based on the HLLC approximate nonlinear Riemann solver for gas dynamics for the ideal MHD equations written in conservation form. The method is intended to be computationally inexpensive (compared to exact nonlinear solvers) and guaranteed in almost all cases to provide positive densities and pressures. The talk includes a brief introduction covering ideal MHD, Riemann problems, and the HLLC method for gas dynamics. Simulation results will concentrate on one-dimensional test cases.

Speaker's web page: http://math.nist.gov/~KGurski

Institution web page: http://www.nist.gov

April 4, 2003

Challenges in the Computational Discovery of Explanatory Scientific Models

Pat Langley
Stanford University & Institute for the Study of Learning and Expertise

Email: langley@csli.stanford.edu

Abstract

The growing amount of scientific data has led to the increased use of computational discovery methods to understand and interpret them. However, most work has relied on knowledge-lean techniques such as clustering and classification learning, which produce descriptive rather than explanatory models, and it has used formalisms developed in AI or statistics, so that results seldom make contact with current theories or scientific notations. In this talk, I present a new approach to computational discovery that encodes explanatory scientific models as sets of quantitative processes, simulates these models' behavior over time, incorporates background knowledge to constrain model construction, and induces these models from time-series data in a robust manner. I illustrate this framework on data and models from earth science and microbiology, two domains in which explanatory process accounts occur frequently. In closing, I describe our progress toward an interactive software environment for the construction, evaluation, and revision of such explanatory scientific models. This talk describes joint work with Kevin Arrigo, Stephen Bay, Lonnie Chrisman, Dileep George, Andrew Pohorille, Javier Sanchez, Dan Shapiro, and Jeff Shrager.

Speaker's web page: http://www.isle.org/~langley/

Research web page: http://cll.stanford.edu/~langley/

Institution web page: http://www-csli.stanford.edu/cll/

March 26, 2003

Existing and Emerging Approaches for Addressing the Processor Memory Gap

Lenny Oliker
Lawrence Berkeley National Laboratories

Email: loliker@lbl.gov

Abstract

The growing gap between sustained and peak performance for scientific applications has become a well-known problem in high-performance computing. This talk presents existing and emerging approaches for addressing this gap. First, we examine the NEC SX6 vector processor, which offers the potential to deliver a substantial increase in computing capabilities for a significant number of computational science codes. We compare the intranode performance between the SX6 vector processor and the cache-based IBM Power3/4 superscalar architectures across a number of key scientific computing areas. We evaluate low-level system characteristics using microbenchmarks and present performance results for numerical codes from scientific computing domains, including astrophysics, fusion energy, materials science, fluid dynamics, and molecular dynamics.

Next we examine two emerging micoarchitectures designed for high-end media processing that have the potential to bridge the processor memory gap. The VIRAM architecture uses novel PIM technology to combine embedded DRAM with a vector coprocessor for exploiting its large bandwidth potential. The Imagine architecture, on the other hand, provides a stream-aware memory hierarchy to support the tremendous processing potential of the SIMD-controlled VLIW clusters. Our experiments isolate the set of application characteristics best suited for each architecture and show a promising direction toward interfacing leading-edge media processor technology with high-end scientific computations.

Speaker's web page: http://www.nersc.gov/~oliker/

Institution web page: http://www.lbl.gov/

March 17, 2003

Middleware Support for Data Ensemble Analysis

Joel Saltz
Ohio State University

Email: saltz-1@medctr.osu.edu

Abstract

Dramatic decreases in the cost of storage, combined with equally dramatic improvements in network connectivity, will allow communities to collaboratively generate and analyze very large distributed datasets. We will describe application scenarios that motivate this work and provide a broad view of what advances in systems software are needed to make this vision a reality. In many application scenarios, datasets describe spatio-temporal regions. We will then describe techniques we have developed to support optimized distributed data storage, indexing, retrieval, and processing. Our approach is to develop systems software able to leverage knowledge of spatio-temporal descriptive metadata in a way that supports a broad range of application areas. We will describe techniques that target very large datasets distributed among storage systems located in multiple parallel machines and clusters. These techniques target spatio-temporal and relational queries directed against these large grid-based datasets. They also target multiple-query optimization techniques that involve identification of intermediate results shared between the queries, along with grid-based sematic caching and retrieval of these results.

Speaker's web page: http://medicine.osu.edu/Informatics/saltz.html

Research web page: http://www.cs.umd.edu/projects/hpsl/chaos

Institution web page: http://www.osu.edu/

March 5, 2003

From Elementary Reactions to Power Plants: In How Much Detail Should Combustion Be Modeled?

Frank Behrendt
Berlin University of Technology

Email: frank.behrendt@tu-berlin.de

Abstract

Modeling and simulation are an established third branch of science in addition to theory and experiment. While modeling and simulation have seen significant successes, principal limitations remain. This talk will address these limitations with respect to combustion processes.

Modeling and simulation of combustion aim at the development of more efficient, cleaner, and more economical combustion technologies. To achieve these developments through modeling, the underlying models have to undergo continuous improvement to increase their predictive power. But even with the multi-teraflops computers available today, only a limited range of scales can be resolved simultaneously. Consequently, to resolve microscopic and macroscopic scales at the same time, hierarchical models are required.

This talk will address hierarchical models for three cases—elementary chemical reaction mechanisms, droplet combustion, and turbulent flows. Principal approaches are discussed and their suitability evaluated. An integrated model for the burning chamber of a coal-fired power plant concludes the talk.

Speaker's web page: http://www.tu-berlin.de/fb6/ifet/rdh/RDH_eng/people.htm

Research web page: http://www.tu-berlin.de/fb6/ifet/rdh/RDH_eng/

Institution web page: http://www.tu-berlin.de/

February 28, 2003

PAPI and Dynamic Performance Analysis Technology

Phil Mucci
University of Tennessee, Knoxville

Email: mucci@cs.utk.edu

Abstract

In this talk, I will cover the status and direction of PAPI, Dynaprof, and related tool technologies. PAPI is an open, cross-platform interface to the performance analysis hardware found in most modern microprocessors. In its simplest form, this hardware consists of a small set of programmable event counters. In addition, it may contain some advanced features such as performance-monitor interrupts, instruction tracing, branch profiling, and address sampling. Monitoring these types of events has a variety of uses in application performance prediction, analysis, and subsequent optimization. PAPI provides two standardized interfaces, one for the application engineer looking for quick measurements and one for the tool developer, both providing maximum performance and flexibility. Dynaprof is a tool to allow the user to insert and remove performance instrumentation (such as PAPI) at runtime. Dynaprof uses either DPCL or DynInst to insert the instrumentation directly into the address space of the application. Dynaprof currently provides a PAPI probe for collecting hardware counter data and a wallclock probe for measuring elapsed time, both on a per-thread basis. Users may write their own probes, and that probe may use whatever output format is appropriate—for example, a real-time data feed to a visualization tool or a static data file dumped to disk at the end of the run. Future plans for Dynaprof will be discussed, including integration with the University of Wisconsin's Tool Daemon Protocol, Lawrence Livermore National Laboratory's Tool Gear System, the University of Oregon's TAU, Sandia National Laboratory's Vprof, and Rice University's PAPIprof.

Speaker's web page: http://www.cs.utk.edu/~mucci

Research web page: http://icl.cs.utk.edu/projects/papi/

Institution web page: http://www.utk.edu

February 19, 2003

A Data Mining Approach to Modeling of Fusion Plasmas

Michael Walker
General Atomics

Email: walker@fusion.gat.com

Abstract

Experimental fusion devices have a useful life of several years and generate enormous quantities of raw and processed data. For example, just the raw diagnostic data archived for the DIII-D fusion experiment now increases by over 400 gigabytes (compressed) per year. Traditional analysis approaches typically focus on a few plasma discharges at a time, sometimes only one, out of the thousands collected per year. An opportunity exists for development and application of software tools that can process large numbers of these discharges to look for patterns in the data, a process usually referred to as data mining.

A significant challenge to developers of scientific data mining technology is the need to provide methods that can derive knowledge from data in a form useful to the scientist within a particular discipline. In tokamak plasma physics, for example, knowledge representations range from experientially derived procedures for how to access various plasma regimes during experimental operations to detailed and sophisticated models of plasma behavior that combine multiple phenomena such as particle transport, effects of external magnetic fields, pressure, or radio-frequency heating. The most sophisticated representations often take the form of detailed simulations, while simpler empirical models, or scaling laws, are often used to provide intuitive understanding.

Empirical physics models are also often the first step in the development of theory derived from observed experimental behavior. This talk will discuss the process of deriving empirical physics models from experimental data when viewed as a computational problem. Overview descriptions of some applications of data mining tools to four representative fusion plasma physics problems will be given.

Institution web page: http://www.generalatomics.com/

February 18, 2003

DEFACTO: Combining Parallelizing Compiler Technology with Hardware Behavioral Synthesis

Pedro Diniz
University of Southern California

Email: pedro@isi.edu

ISI Institute for Scientific Computing Research

Abstract

Field Programmable Gate Arrays (FPGAs) offer the promise of substantial performance improvements over conventional processors by allowing the implementation of application-specific data paths that exploit instruction-level parallelism or domain-specific numeric formats and operations. Unfortunately, FPGAs are still difficult to program, making them inaccessible to the average developer. The standard practice requires developers to express the application program in a hardware-oriented language such as Verilog or VHDL and to synthesize the hardware design using a variety of synthesis tools. Because of the complexity of the synthesis process, it is difficult to predict a priori the performance and space characteristics of the resulting design. For this reason, the developer usually engages in an iterative design process, examining the results and modifying the design to trade off performance for space.

In this presentation, we describe DEFACTO, a system that automatically maps computations written in high-level imperative programming languages such as C to multi-FPGA-based systems. DEFACTO combines parallelizing compiler technology with commercially available behavioral synthesis tools. We use synthesis estimation techniques to guide the application of high-level program transformations in the search of high-quality hardware designs, thereby avoiding the long compilation–synthesis design cycles. We illustrate the effectiveness of DEFACTO in exploring a wide space of implementation designs for a set of image processing computations. For these computations the system searches on average less than 0.3% of the design space while deriving an "optimal" implementation, leading up to four orders of magnitude reduction in design time. We also describe in detail the mapping of a digital image processing computation, the Sobel edge detection, for which DEFACTO yields a 60-fold reduction in design time with only a 59% increase in execution time as compared to a manual implementation of the same algorithm.

This work has been sponsored by the Defense Advanced Research Project Agency (DARPA) under contract # F30602-98-2-0113.

Speaker's web page: http://www.isi.edu/~pedro/

Research web page: http://www.isi.edu/asd/defacto

Institution web page: http://www.usc.edu

February 18, 2003

Computational Approaches Towards Functional Annotation of the Human Genome

Ivan Ovcharenko

Lawrence Berkeley National Laboratory

Email: ovcharenko1@llnl.gov

Abstract

A first glance at the complete genomic sequence of humans revealed evidence for about 30,000 unique genes. The proteins encoded by these genes are responsible for constructing the human body from a single cell and regulating all the biochemical processes required for maintaining human health. While there is extensive knowledge about the function of many human genes from studies in model organisms and from hints from genetic disorders, the sequences required for the transcriptional regulation of these genes are poorly defined or as of yet unidentified. For example, why is BMP4, a gene responsible for teeth generation, shared by humans and birds, but the latter lack teeth? What sequences differentially regulate the expression pattern of BMP4 in birds and humans? How can we use computational tools to assist us in identifying regulatory DNA sequences encoded in the human genome?

Comparative sequence analysis has been proven to be an efficient and powerful approach to identify functional regions in the human genome. Genomes of different organisms have significantly diverged throughout the evolution of life on Earth, while orthologous genes and functional regulatory elements have been evolutionarily conserved. This talk will introduce computational strategies for aligning the human, mouse, and fugu genomes, and for predicting putative regulatory elements by combining transcription factor binding site prediction and the analysis of inter-species sequence conservation. I will also discuss how we can use the information derived from sequence alignments to understand the landscape of the human genome in terms of gene density, evolutionary conservation, and noncoding gene regulation.

February 14, 2003

Efficient Discovery of Previously Unknown Patterns and Relationships in Massive Time Series Databases

Eamonn Keogh
University of California at Riverside

Email: eamonn@cs.ucr.edu

Abstract

To date, the vast majority of research on time series data mining has focused on similarity search and, to a lesser extent, on clustering. We believe that these problems should now be regarded as essentially solved. From a knowledge-discovery viewpoint, several important unsolved problems in time-series data mining are more interesting, important, and challenging. In future work, we intend to address these problems. Our long-term goal is the creation of efficient algorithms to allow the extraction of knowledge in the form of patterns, anomalies, regularities, and rules from massive time-series datasets.

We will begin with a summary of the state of the art in time-series data mining, including a review of the multitude of representations proposed in the literature. We will then concretely define the three major problems we wish to address:

- enumerating frequently occurring patterns in time-series databases,
- the discovery of surprising or anomalous patterns in time series, and
- the discovery of causal/association rules in time series.

We will consider the limitations of the current work in these areas, enumerate the challenges faced, and sketch out a promising unpublished idea with implications for all these problems.

Speaker's web page: http://www.cs.ucr.edu/~eamonn/

Institution web page: http://www.cs.ucr.edu

February 7, 2003

Mixed Finite Elements for Elasticity

Douglas Arnold
University of Minnesota

Email: arnold@ima.umn.edu

Abstract

There have been many efforts, dating back four decades, to develop stable mixed finite elements for the stress-displacement formulation of the plane elasticity system. The key is to design a finite-element discretization for the space of square-integrable symmetric tensor fields with square-integrable divergence. This discretization enjoys a variety of properties. Although there are a number of well-known discretizations of H(div) vector fields with the analogous properties, such finite elements for symmetric tensor fields have proven very hard to design. We will present a new family of such elements, one for each polynomial degree quadratic and above. We will also analyze the obstructions to the construction of such elements, which account for the paucity of elements available. A star supporting role will be played by the de Rham sequence and related sequences of partial differential operators and their discrete analogues.

Speaker's web page: http://www.ima.umn.edu/~arnold/

Institution web page: http://www.umn.edu/

February 6, 2003

The Institute for Mathematics and Its Applications

Douglas Arnold
University of Minnesota

Email: arnold@ima.umn.edu

ISCR Institute for Scientific Computing Research

Abstract

The Institute for Mathematics and Its Applications is one of the world's foremost institutions devoted to increasing the impact of mathematics by fostering interdisciplinary work. In this informal presentation, the director of the Institute for Mathematics and Its Applications will give an overview of upcoming activities and opportunities at the IMA.

Speaker's web page: http://www.ima.umn.edu/~arnold/

Institution web page: http://www.umn.edu/

February 3, 2003

Software Quality Process: Is It a Square Peg in a Round Hole?

Stan Rifkin
Master Systems

Email: sr@Master-Systems.com

Abstract

We in the business of helping organizations change often label the members of our audience "bad characters" because they don't agree with us. This presentation offers a different view: our clientele are not bad characters, and if we listened to them and acted on what is underneath what they say, the problem of implementation transforms from convincing, persuading, and cajoling to controlling wild horses that want to dash off with the new initiatives.

The key observation is that our software process improvement programs are not aligned with the strategy—stated or not—of our organizations, and those whom we label as resisters are actually performing a valuable service by pointing out the gaps between what drives the organization and what we are proposing.

There are only three strategies, and, oddly enough, many quality improvement agendas are not the same as what gets rewarded, so there is a natural tension between our proposals and what the organization does. This presentation offers concrete antidotes, surprisingly not by changing the rewards for quality but rather by changing the quality programs.

Speaker's web page: http://www.master-systems.com/Papers.ivnu

Institution web page: http://www.master-systems.com/

January 13, 2003

Geometric Programming with a Functional Language

Alberto Paoluzzi
Third University of Rome, Italy

Email: paoluzzi@dia.uniroma3.it

Abstract

The aim of this talk is to introduce the new features of PlaSM, a design environment for geometric modeling and animation that supports rapid prototyping of geometric structures but does not deprive the user of direct control over the underlying geometric programming. PLaSM is a geometry-oriented extension of a subset of FL, developed by the Backus Functional Programming Group at IBM Almaden, that introduced an algebra over programs, in which a set of algebraic identities between functional expressions is established. PLaSM was recently extended with object-oriented features and with heuristic functions, allowing for some rule-based forward-chaining reasoning. The approach to the integration of the three paradigms, to our knowledge, is new. Joint work with CASC researchers is under way to design a new geometric kernel based on progressive combination and visualization of large-scale spatial indices. The PLaSM project (http://www.plasm.net) is opensource and multiplatform (windows, linux, macosx). A book on the subject is in press.

Speaker's web page: http://www.dia.uniroma3.it/~paoluzzi/

Research web page: http://www.plasm.net/

Institution web page: http://www.dia.uniroma3.it/index-en.html

January 10, 2003

Active Learning with Multiple Views

Ion Muslea
University of Southern California

Email: muslea@isi.edu

Abstract

Labeling training data for machine learning algorithms is tedious, time-consuming, and error-prone. Consequently, it is of utmost importance to minimize the amount of labeled data that is required to learn a target concept. In the work presented here, I focus on reducing the need for labeled data in multiview learning tasks. The key characteristic of multiview learning tasks is that the target concept can be independently learned within different views (i.e., disjoint sets of features that are sufficient to learn the concept of interest). For instance, robot navigation is a two-view learning task because a robot can learn to avoid obstacles based on either sonar or vision sensors. In my dissertation, I make three main contributions. First, I introduce Co-Testing, which is an active learning algorithm that exploits multiple views. Co-Testing is based on the idea of learning from mistakes. More precisely, it queries examples on which the views predict a different label: if two views disagree, one of them is guaranteed to make a mistake. In a variety of real-world domains, from information extraction to text classification and discourse-tree parsing, Co-Testing outperforms existing active learners. Second, I show that existing multiview learners can perform unreliably if the views are incompatible or correlated. To cope with this problem, I introduce a robust multiview learner, Co-EMT, which interleaves semisupervised and active multiview learning. My empirical results show that Co-EMT outperforms existing multiview learners on a wide variety of learning tasks. Third, I introduce a view validation algorithm that predicts whether or not two views are adequate for solving a new, unseen learning task. View validation uses information acquired while solving several exemplar learning tasks to train a classifier that discriminates between tasks for which the views are adequate and inadequate for multiview learning. My experiments on wrapper induction and text classification show that view validation requires a modest amount of training data to make high-accuracy predictions.

Speaker's web page: http://www.isi.edu/~muslea/

Research web page: http://www.isi.edu/~muslea/papers.html

Institution web page: http://www.isi.edu

January 6, 2003

Computational Problems in Proteomics: Statistics, Optimization, and Combinatorics

Alex Pothen
Old Dominion University

Email: pothen@cs.odu.edu

ISCR Institute for Scientific Computing Research

Abstract

Now that the genomes of many organisms have been sequenced, large-scale projects are under way to characterize the protein products of the genes (the proteome) and the multiprotein complexes that are responsible for the functions of the cell. High throughput, rapid, and automatable techniques are currently being developed to identify the tens of thousands of proteins (or more) involved. Among these are protein chips, novel mass spectrometric techniques, and nuclear magnetic resonance (NMR) methods.

Two computational problems in proteomics will be considered in this talk.

The first problem is to use the protein profiles of tissues, obtained via novel mass spectrometric techniques (SELDI and MALDI), to classify tissues into diseased or healthy specimens. Statistical classification and support vector machines are used to discover protein markers that characterize disease.

The second problem is to computationally represent multiprotein complexes and protein interaction networks using graphs and hypergraphs to enable algorithms for answering biological questions. We describe a new algorithm for identifying a "k-core" of a hypergraph (a subhypergraph in which every vertex belongs to at least "k" hyperedges of the subhypergraph) and use it to characterize core proteomes of yeast. The biological significance of core proteomes is that they are expected to have similar functions in related organisms.

Speaker's web page: www.cs.odu.edu/~pothen

Institution web page: www.cs.odu.edu/

December 13, 2002

Digital Libraries for Scientific Data Management

Reagan Moore
San Diego Supercomputer Center

Email: moore@sdsc.edu

Abstract

Across the data-grid, digital-library, and persistent-archive communities, a consensus is emerging on the software infrastructure needed for scientific data management. Each community has focused on a particular set of capabilities needed to manage data distributed across administration domains and multiple types of storage repositories. Data grids look at latency management and storage repository federation. Digital libraries look at discovery mechanisms for finding relevant material. Persistent archives examine the management of technology evolution. The common approach that is emerging is based upon:

- the use of logical name spaces to provide infrastructure-independent naming,
- storage repository abstractions to define the set of operations used to manipulate remote digital entities, and
- information repository abstractions to define the set of operations needed to manipulate collections stored in databases.

The common approach will be discussed, along with representative projects from each of these data-management communities.

Research web page: http://www.sdsc.edu/~moore/rmoore.html

Institution web page: http://www.sdsc.edu/

December 11, 2002

Tools and Benchmarks for Performance Evaluation of Applications in Science

Rudolf Eigenmann
Purdue University

Email: eigenman@ecn.purdue.edu

Abstract

I will give an overview of a number of projects that create programming environments and a performance evaluation infrastructure for computational applications. Two important parts of the environment are the Ursa Minor interactive performance tuning tool and the Polaris parallelizing compiler. Ursa Minor combines information gathered from sources such as compilers and runtime performance monitors and presents it in a uniform way to the user. It supports stepwise performance improvements, following a methodology of performance analysis and improvement for OpenMP programs. The Polaris compiler furnishes Ursa Minor with information about program characteristics and automatable parallelism. An important part of every performance evaluation infrastructure is the test program suite, i.e., benchmarks. We have participated in an effort to collect the largest possible public computational applications and to create benchmark suites for both research and industry. The most recent suite is just being released under the name SPEC HPC2002.

Speaker's web page: http://ece.www.ecn.purdue.edu/~eigenman/Index.html

Research web page: http://ParaMount.www.ecn.purdue.edu/ParaMount/

Institution web page: http://ece.www.ecn.purdue.edu

December 6, 2002

The Oxymoron of Computer "Science"

John Perry
Foothill Deanza College

Email: perry@voyager.deanza.fhda.edu

Abstract

The author gives a historical account of the field of computer science since the 1970s. The purpose of this historical review is to show how the field is, in fact, not a science at all but rather the practicing of pop psychology by charismatic Ph.D.'s, the imposition of empirically unproven paradigms and procedures on programmers, and a sociological phenomenon replete with elements of "hipness" and associated in-group and outcast behaviors. The author illustrates how the celebritization of personalities, the religification of computers, and the lack of respect for true science have led to regressive software, regressive programming languages, regressive paradigmatic systems, and the death of common sense in a field allegedly allied with the "hard" sciences.

Institution web page: http://www.deanza.edu/

December 3, 2002

Detecting and Exploiting Spatial Regularity in Memory References

Tushar Mohan
University of Utah

Email: tushar@cs.utah.edu

Abstract

Rising processor speed, unaccompanied by corresponding reductions in memory access latency, has caused the performance of codes to be limited by memory accesses. Strided memory accesses, or streams, if known to exist in an application, can be targeted by a host of optimizations, such as stream prefetching, relocation, remapping, and vector loads. Undetected, they can be a significant source of memory stalls in loops. Existing stream-detection mechanisms either require special hardware, which may not gather enough stream statistics for subsequent analysis, or are confined to limited compile-time detection of array accesses in loops. Little formal treatment has been accorded to the subject; the concept of locality fails to capture the existence of streams in a program's memory accesses. In this thesis we define spatial regularity as a means to depict the presence of strided memory accesses. We develop measures to quantify spatial regularity and design and implement an online, parallel algorithm to detect streams, and hence regularity, in running applications. We identify critical program sections for regularity measurements by using PAPI—a performance measurement API—to access hardware performance counters portably. Dyninst's dynamic binary translation infrastructure is leveraged to perform selective and transitory instrumentation in the application. This development allows the user to limit the stream detection overhead at the cost of measurement accuracy. We use examples from real codes and popular benchmarks to illustrate how stream information can be used to effect profile-driven optimizations.

Speaker's web page: http://www.cs.utah.edu/~tushar/

Institution web page: http://www.utah.edu/

November 8, 2002

Link-Time Optimization of Parallel Scientific Programs

Gregory Andrews
University of Arizona

Email: greg@CS.Arizona.EDU

Abstract

A binary rewriting system transforms a binary program into a different but functionally equivalent program. A link-time optimizer is a binary rewriting system that optimizes an object program to improve some aspect of its performance, such as execution time, code size, power consumption, or communication bandwidth. A significant benefit of a link-time optimizer is that it can perform whole program optimizations that cannot be done by compilers—e.g., those based on run-time values—or that usually are not done by compilers—e.g., because library source code is not available or is written in a different language.

The SOLAR project is developing binary rewriting techniques for software optimization at link time and run time. A particular interest is improving the performance of parallel scientific programs that use a communications library such as MPI. This talk will describe the link-time optimizer we have developed for the Pentium architecture and demonstrate the improvements it is able to make to already highly optimized application programs. The talk will also describe several useful tools that have been relatively easy to construct given our Pentium link-time optimizer.

Speaker's web page: http://www.cs.arizona.edu/people/greg/

Institution web page: http://www.arizona.edu/

October 29, 2002

On Balanced Approximations for Time Integration of Multiple Time Scale Systems

Dana Knoll
Los Alamos National Laboratory

Email: nol@lanl.gov

ISCR Institute for Scientific Computing Research

Abstract

The effect of various numerical approximations used to solve linear and nonlinear problems with multiple time scales is studied in the framework of modified equation analysis (MEA). MEA is used to study the effect of linearization and splitting in a simple nonlinear ordinary differential equation (ODE) and in a linear partial differential equation (PDE). Several forms of time differencings of the ODE and PDE are considered, and the resulting truncation terms are compared analytically and numerically. It is demonstrated quantitatively that both linearization and splitting can result in accuracy degradation when a computational time step larger than the competing (fast) time scales is employed. Many of the issues uncovered on the simple problems are shown to persist in more realistic applications.

Institution web page: http://www.lanl.gov

October 28, 2002

Physics-Based Preconditiong for Jacobian-Free Newton–Krylov Methods

Dana Knoll
Los Alamos National Laboratory

Email: nol@lanl.gov

Abstract

An important class of preconditioners for the Jacobian-free Newton–Krylov method is referred to as physics-based or PDE-based. The motivation behind this approach is that there exist numerous, legacy algorithms to solve nonlinear systems, both IVPs and BVPs. These algorithms typically were developed with some insight into the time scales or physical behavior of the problem. As a benefit of this insight, a reduced implicit system, or a sequence of segregated explicit or implicit systems, may be solved in place of the fully coupled system. Examples include the semi-implicit method for low-speed flow (stiff-wave problem), the SIMPLE algorithm for incompressible flow, Gummel's method for the semiconductor drift-diffusion equations, and numerous other structure-based operator splitting methods for reaction-diffusion systems. We will outline the concept of physics-based preconditioning and then focus on our own work on applications with stiff-wave phenomena.

Institution web page: http://www.lanl.gov/

October 18, 2002

Numerical Study of Burn Propagation in Inhomogeneous Mixtures

David Lopez and
Sutanu Sarkar
University of California, San Diego

Email: ssarkar@ucsd.edu

Abstract

In inertial confinement fusion applications, the reactant mix can be contaminated by inert material introduced by Rayleigh–Taylor hydrodynamic instabilities promoted at sites of imperfect surface finish of the capsule. We have performed multidimensional simulations to study burn-front propagation through a contaminated mixture. The flow field and chemistry are accurately resolved on an instantaneous space–time basis. Associated with the large density change across the burn front and consequent fluid acceleration, there is a natural instability, the so-called Darrieus–Landau instability discussed in premixed combustion literature. The coupling of contaminant and velocity fluctuations in the burn mix with the natural instability is found to significantly affect burn-front propagation. In the talk, we will discuss the simulation method and results illustrating the modified burn propagation in a contaminated mix.

Research web page: http://www-mae.ucsd.edu/RESEARCH/SARKAR/sarkar.html

Institution web page: http://www.ucsd.edu/

October 8, 2002

Learning to Classify Galaxy Shapes Using the EM Algorithm

Sergey Kirshner
University of California, Irvine

Email: skirshne@odessa.ics.uci.edu

Abstract

I will describe the application of probabilistic model-based learning to the problem of automatically identifying classes of galaxies, based on both morphological and pixel intensity characteristics. The EM algorithm can be used to learn how to spatially orient a set of galaxies so that they are geometrically aligned. This "orientation-model" is augmented with a mixture model on objects, and classes of galaxies can then be learned in an unsupervised manner using a two-level EM algorithm. The resulting models provide highly accurate classification of galaxies in cross-validation experiments.

Research web page: http://www.ics.uci.edu/~skirshne/

Institution web page: http://www.uci.edu/

October 3, 2002

Leveraging PC Graphics Cards for Advanced Visualization

Charles Hansen
University of Utah

Email: hansen@cs.utah.edu

Abstract

Recent advances in the programmability of PC graphics hardware have led to corresponding advances in direct volume rendering. In this talk, I will present an overview of two such advances: the use of multidimensional transfer functions and the incorporation of advanced shading techniques. Most direct volume renderings produced today employ one-dimensional transfer functions, which assign color and opacity to the volume based solely on the single scalar quantity that comprises the dataset. Though they have not received widespread attention, multidimensional transfer functions are a very effective way to extract specific material boundaries and convey subtle surface properties. However, identifying good transfer functions is difficult enough in one dimension, let alone two or three. I will outline an important class of three-dimensional transfer functions for scalar data (based on data value, gradient magnitude, and a second directional derivative), and describe a set of direct manipulation widgets that make setting such transfer functions intuitive and convenient. I will also describe how to use modern graphics hardware to interactively render with multidimensional transfer functions. The transfer functions, widgets, and hardware combine to form a powerful system for interactive volume exploration.

Direct volume rendering is a commonly used technique in visualization applications. Many of these applications require sophisticated shading models to capture subtle lighting effects and characteristics of volumetric data and materials. Many common objects and natural phenomena exhibit visual quality that cannot be captured using simple lighting models or cannot be solved at interactive rates using more sophisticated methods. I will describe a simple yet effective interactive shading model, which captures volumetric light attenuation effects to produce volumetric shadows and the subtle appearance of translucency. By coupling this shading with a technique for volume displacement or perturbation, one can achieve realistic interactive modeling of high-frequency detail for real and synthetic volumetric data.

Research web page: http://www.cs.utah.edu/~hansen/

Institution web page: http://www.utah.edu/

Institute for Scientific Computing Research

ISCR Summer Student Program

ISCR Summer Student Program

Each summer the ISCR runs an extensive summer program for students ranging from those still in high school to those completing their PhDs. Students are assigned specific projects appropriate to their research capabilities. A technical mentor oversees the work and provides guidance. While in residence, students attend research seminars, prepare a poster summary of their work, and write up a progress report on their efforts.

Time-Critical Occlusion Culling: A Multiresolution Approach

Lucas Ackerman
Worcester Polytechnic Institute
Mentor
Mark Duchaineau
CASC

Summary

This research is the result of an ongoing three-year collaboration with Mark Duchaineau on scalable visibility algorithms.

We have developed an algorithm for conservative visibility determination on multiresolution surfaces, suitable for complementing continuous level-of-detail (LOD) rendering algorithms. Fundamental issues include the feedback between LOD and visibility requirements, the methods of performing occlusion with elements of multiresolution geometry, and the unique benefits of calculating visibility in an inherently multiresolution domain. This is the only approach known to be adaptive in object space on both occludee and occluder hierarchies. It requires no precomputation beyond the nested error bounds used for continuous LOD display (as in the ROAM algorithm) and is amenable to being made progressive and output-sensitive with respect to the computed visible set.

This algorithm complements the ROAM family of continuous LOD algorithms.

Future plans include alternate implementations to improve performance, address shortcomings, and increase flexibility, and a formal paper for publication.

Multigrid for Maxwell's Equations on Structured and Unstructured Grids

David Alber
University of Illinois, Urbana–Champaign

Mentor

Jim Jones
CASC

Summary

Maxwell's equations unify the concepts of the electric and magnetic fields. As such, they are fundamental to electromagnetics and constitute an important problem to discretize and solve numerically with good accuracy and speed. Unfortunately, the discrete Maxwell's problem is not solved well by standard methods because of the operator of the discretized problem. The Maxwell operator has components that are near null-space but are not smooth. This means that standard relaxation will not damp this error component because it is algebraically smooth, and coarse grid correction will not remove the error because it cannot be represented on a coarse grid. This difficulty suggests that new smoothers, new grid-transfer operators, or both need to be developed.

Both the definite and indefinite formulations of Maxwell's equations were examined this summer. In the definite formulation, the near null-space components of the operator were handled by an overlapping Schwarz smoother, as suggested by Arnold, Falk, and Winther. Good rates for this problem on a structured grid were achieved using this smoother and standard coarsening on the unknowns. A solver for the definite Maxwell system on unstructured grids was also developed. This solver uses element agglomeration techniques to select the coarse grid and also uses the overlapping Schwarz smoother. The rates that have been observed for this solver are as good as the rates from the solver on the structured grid. However, it is premature to draw too many conclusions from this result because of limited testing. More work needs to be done on this solver. The indefinite formulation of Maxwell's equations offers additional complications in the form of near null-space plane waves. These waves are difficult to eliminate, and thus far, a good solver has not been found for this problem.

This project will continue with further work on solvers for both definite Maxwell's on unstructured grids and the indefinite Maxwell's formulation. Many pieces of the solver on the unstructured grid will be examined, including the agglomeration routines, the smoother, and the prolongation operator.

Software Quality Assurance and Miscellaneous Changes in DRACO

John C. Anderson
University of the Pacific
Mentor
Benjamin T. Grover
DCOM

Summary

In large software projects, the ability to quantitatively assess both correctness and quality is important. It is impractical to employ traditional quality control techniques such as peer-review when dealing with a large, rapidly changing code base. DRACO, a parallel mesh generation tool with over 275,000 lines of code, was a prime candidate for two modern methods of quality assurance: Design By Contract (DBC) programming and unit testing [2]. The goal of the investigation was to implement software quality tools for DRACO addressing the need for DBC as well as unit testing.

Over the summer, I was able to implement a complete set of tools for software quality assurance in DRACO. These tools include two DBC systems, one for C++ and another for Python, an extended testing framework complete with over 175 unit tests and more testing modes, and a robust C++ debug flag system with an easy-to-use Python interface.

After finishing the above software quality enhancements, I worked to improve and extend other areas of DRACO. Improvements were made to the build system, installation script, the GUI, and the handling of material attributes on meshes. Extensions included the development of a function browser for executing DRACO commands from the GUI, a Boolean Operations dialog, and the addition of a flexible view mode system for DRACO's graphical display.

Performance of I/O Nodes in the BlueGene/L Supercomputer

Scott Banachowski
University of California, Santa Cruz
Mentor
Kim Yates
CAR

Summary

The BlueGene/L supercomputer (BG/L) will be a massively parallel system of 65,536 compute nodes. The system software of BG/L is architecturally organized into 1024 logical groups of 64 compute processors. To offload many operating system tasks such as file I/O, sockets, and debugger services from compute nodes, each logical group employs an additional I/O node that runs an embedded version of Linux, whereas compute nodes run a small, simple, fast kernel. The compute nodes ship I/O requests, such as file reads and writes and other system operations, to their I/O node for execution. The goal of the project is to investigate if an I/O node has the computational power and resources to support 64 compute nodes without performance bottlenecks.

A suite of benchmarking software was developed for this project. It consists of two applications: a parallel application, which runs on the compute nodes and creates load to stress the I/O node by making bulk file operations, and a monitor application, which runs on the I/O node and measures statistics about the node's resource usage, including memory, network bandwidth, and processor idleness. The system software that runs on the I/O node was not available to view or modify, so we must determine what is happening in the I/O node through other means. The I/O node runs a version of Linux and, unlike the compute nodes, is capable of multitasking. Therefore, it is possible to launch an application that makes resource measurements on the I/O node during the execution of the load experiments. The I/O node monitoring application uses facilities that Linux provides for system status through the proc file system (a pseudo-file system that supports a filelike interface to kernel data structures). During the summer, the benchmarking applications were developed and tested on the BG/L simulator. The simulator was useful for application debugging, but because it is not timing-accurate, any performance inferred from it is not representative of actual performance. In the final week of the practicum, the tests were executed on a 128-node BG/L prototype at IBM Watson Research Center. Although many of the features we wished to test were not entirely implemented, we were able to capture some data from real hardware. Our data shows that the I/O nodes appear capable of scaling to the task of handling the system requests of 64 compute nodes.

Because the BG/L system is still in development, it is a bit premature to measure its performance. For example, the file system prototype is not yet performing near the specified level, so it created an artificial bottleneck in the measured performance. In the future, when the system is more mature, the software developed in this project may be used to further study the interaction between BG/L compute and I/O nodes.

Continued

Scott Banachowski
University of California, Santa Cruz

Summary continued

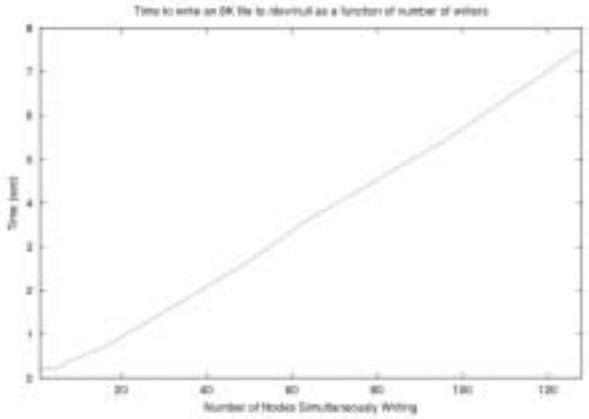


Fig. 1: We varied the number of nodes concurrently writing data to determine how long processing takes when each node writes 8 Kb of file data. To remove the effect of file system latency and measure only operating system overhead, all the data are directed to a null file.

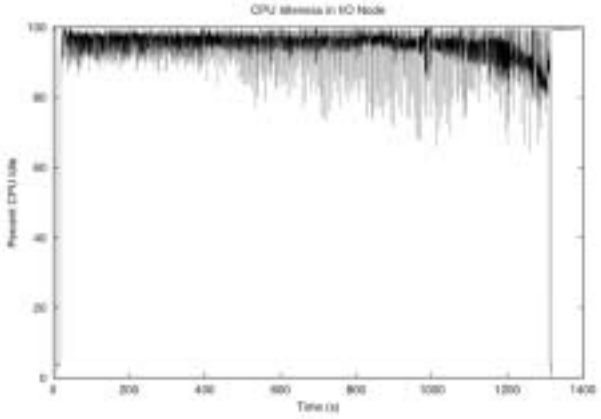


Fig. 2: This is the amount of user and system CPU remaining idle when 64 nodes simultaneously write a file.

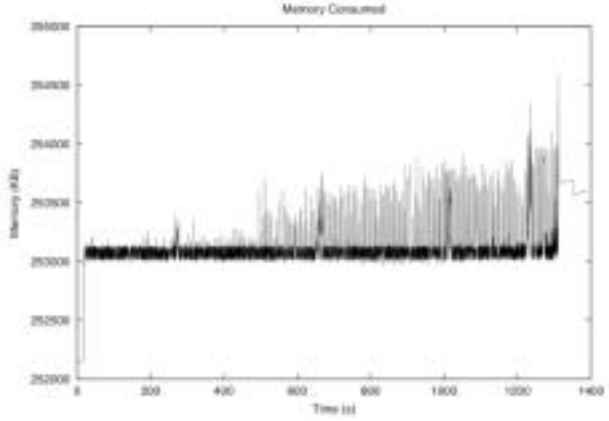


Fig. 3: This plot shows the memory consumed by the I/O node when 64 nodes write a file.

Slow-Growing Subdivision for Generic Unstructured Meshes

Janine Bennett
University of California, Davis
Mentor
Valerio Pascucci
CASC

Summary

Scientific visualization techniques strive to maintain stringent error bounds to ensure that the visual feedback provided to the user or scientist is accurate and reliable. Scientific simulations often use unstructured mesh domains of arbitrary topology with data values sampled at the mesh nodes. However, a hierarchical, regular mesh structure is preferred for efficient, multiresolution access, computation, analysis, and rendering of the data. As a result, the original meshes are often modified for visualization or other postprocessing purposes using interpolation schemes that differ from those used by the simulation. The resulting output is inherently inaccurate, demonstrating the fundamental inefficiency of current representation schemes used for unstructured meshes.

Subdivision methods are one of the most successfully used techniques in multiresolution data representation for surface meshes. Initial research has been done to extend these techniques to the multiresolution representation of volumetric data. For example, the slow-growing subdivision (SGS), first described by Pascucci, is a subdivision method that generalizes to volumetric and higher dimensional meshes. We started with a basic implementation of this approach for convex domains and regular curvilinear grids, where a mapping to a regular structure is provided explicitly (see figure). In this initial stage, we have successfully provided new wavelet representations that can be used to represent associated field data.

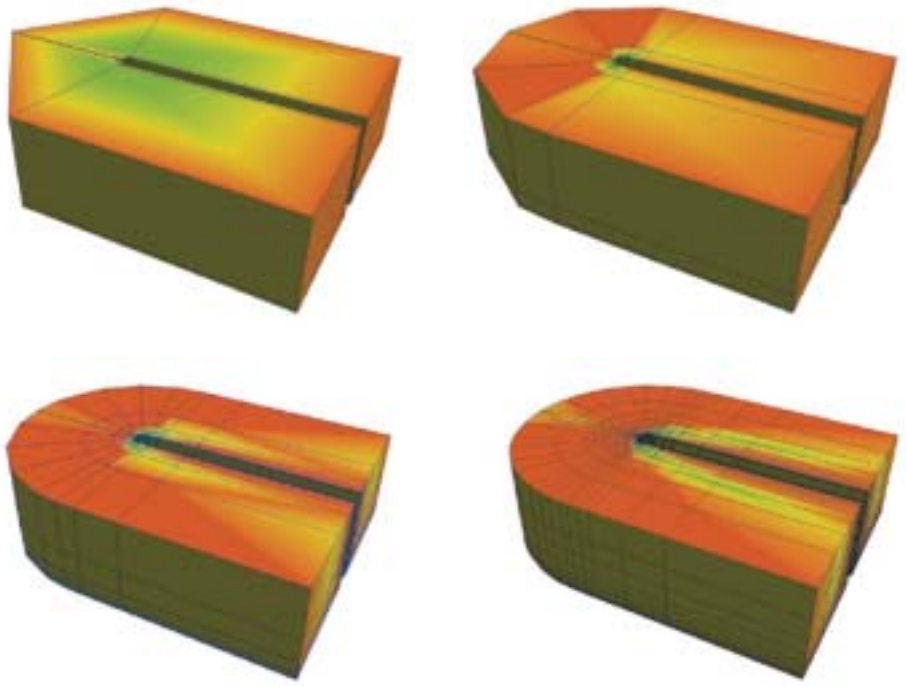
Our main focus for the summer was to complete the postprocessing, compression, and visualization process for SGS representation of arbitrary unstructured meshes. To accomplish this, we decompose the input domain into regions with convexlike connectivity. To this end, we are experimenting with a region growing technique that starts from single-cell sets that are expanded as long as their structural properties are maintained. After this process, we map each region to a convex domain to obtain the necessary isomorphism from unstructured mesh connectivity to subdivision connectivity.

While the completion of a method to remesh a generic unstructured mesh to a mesh on which SGS could be performed is a high priority and is important for visualization purposes, our long-term goal is to provide a unified representation of unstructured mesh data that can be used in simulations and visualization techniques. This data representation would not only produce final images that are inherently more accurate, but could also be used to make simulations more efficient. To this end, we need to work on two major challenges: analyze the formal asymptotic properties and the practical behavior of the nested function spaces associated with the subdivision process and determine how to maintain mesh quality metrics while performing the recursive subdivision process.

Continued

Janine Bennett
University of California Davis

Summary continued



SGS refinement of a nonrectilinear mesh (blunt fin from NASA) with temperature field visualized by pseudocolor map.

Network Vulnerability Assessment (NVA) Project

Bridget Benson
California Polytechnic State University

Mentor

Terry Brugger
NAIC

Summary

The goal of the network vulnerability assessment (NVA) project is to allow computer security personnel to easily perform a network vulnerability assessment. The NVA project involves developing and using tools to collect as much data as possible about a particular network and then displaying this data in a graphical user interface (GUI) to allow computer security personnel and penetration testers to understand the network and look for vulnerabilities.

The NVA project's GUI, known as Graph Viewer, needed bug fixes and enhancements to make it a more useful tool for finding vulnerabilities in the network. Some of these enhancements involved adding color keys and extending the viewer's capability with grouping nodes in the network.

To make enhancements to Graph Viewer, I had to become familiar with its large, existing Java code base. I used the integrated development environment known as Eclipse to step through thousands of lines of code to understand how the viewer was getting, drawing, and arranging the network data.

With the help of my mentor, Terry Brugger, and a new colleague, Chris Brand, with experience in Java Swing and Graphics classes, I was able to make many improvements to Graph Viewer including:

- adding the ability to color network connections based on some information about that connection,
- adding color keys for all color coding done in the viewer,
- fixing a bug to make sure layouts involving groups of computers worked correctly,
- coloring groups of machines,
- extending the expand and reduce functions,
- supporting multiple routes to hosts,
- making labeling by hostname more useful,
- changing the icon of a node in the network based on what server it represents, and
- adding support for analyzing network topology over time.

Through making improvements to Graph Viewer, I acquired a better understanding of the types of vulnerabilities a computer analyst might look for in a network and a great deal of debugging and coding (especially with Java Swing) experience.

The improvements made to the Graph Viewer will aid in meeting the overall objectives of the NVA project. Computer security professionals will continue to make enhancements on Graph Viewer and all other parts of the NVA project every day. With the continual improvements to the NVA project, computer security professionals at Lawrence Livermore National Laboratory are approaching their vision of an automated, comprehensive network vulnerability assessment system.

Massive Volumetric Visualization through Progressive Algorithms and Data Structures

Rita Stefania Borgo
Universita' di Pisa
Mentor
Valerio Pascucci
CASC

Summary

Projects dealing with massive amounts of data need to carefully consider all aspects of data acquisition, storage, retrieval, and navigation. The recent growth in size of large simulation datasets still surpasses the combined advances in hardware infrastructure and processing algorithms for scientific visualization. The cost of storing and visualizing such datasets is prohibitive, so that only one out of every hundred time-steps can be stored and visualized. As a consequence, interactive visualization of result is going to become increasingly difficult, especially as a daily routine from a desktop. High frequency of I/O operations starts dominating the overall running time. In this panorama, the efficiency of a visualization algorithm must be evaluated in the context of end-to-end systems instead of being optimized individually.

Starting from the point that relying on memory layout to speed up the access time is no longer possible because datasets are too large to be kept in main memory or stored on a local disk, a need arises at system level to design the visualization process as a pipeline of modules able to process data in stages, thus creating a flow of data that need to be optimized globally with respect to magnitude and location of available resources.

My objective during the time spent at Lawrence Livermore National Laboratory has been to develop a new progressive visualization algorithm based on edge-bisection refinement, a technique widely used in the meshing community.

The first part of my project was to implement and apply the edge-bisection technique in a 3D environment, with improving visualization of 3D datasets as the overarching goal. Edge-bisection has been implemented based on a set of simple rules that characterize consistently the decomposition of a grid in simplices together with the recursive refinement of the derived simplicial mesh. The result is a new naming scheme that allows representation of an adaptive simplicial mesh with a very low memory footprint.

One property of the subdivision scheme is that it holds implicitly a hierarchical organization of the dataset itself. It has therefore been possible to traverse and organize the input grid in a hierarchical structure (from coarse level to fine level) and, as a consequence, to extract subsequent level of detail for improving display of the output image. As a first step, I have uncoupled data extraction from its display, splitting the overall process between two main threads, one that traverses the input 3D mesh and builds up the hierarchy, and a second that performs traversal of the hierarchy and display of the extracted level of isosurface. This approach allows rendering at any given time partial results while the computation of the complete hierarchy makes progress. The regularity of the hierarchy allows the creation of a good data-partitioning scheme that allows balancing processing time and data migration time. Some results of the progressive behavior of the algorithm are shown in Figure 1.

Continued

Rita Stefania Borgo
Universita' di Pisa

Summary continued

Results in this field are applicable in parallel and distributed computing ranging from a cluster of PCs to more complex and expensive architectures. My next step will be to extend the framework built so far to a shared and distributed memory environment.

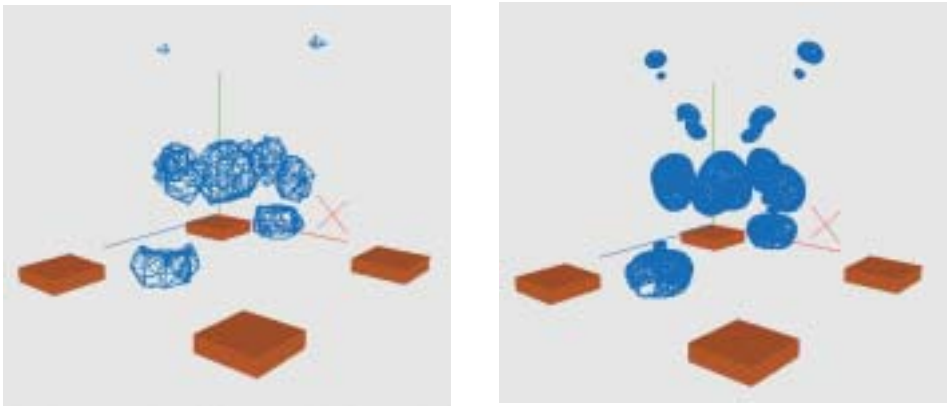


Fig. 1. Illustration of stages of the visualization hierarchy.

Compatible Relaxation in Generalized Algebraic Multigrid (AMG)

James Brannick
University of Colorado, Boulder
Mentor
Rob Falgout
CASC

Summary

In generalizing the algebraic multigrid (AMG) framework, Rob Falgout and Panayot Vassilevski prove that compatible relaxation (CR) can be used to measure the ability of the coarse grid to represent algebraically smooth error. In particular they show that a sufficient condition for guaranteeing the existence of P, the prolongation operator, is that CR is fast to converge. However, the question of how to form such a coarse variable set remains open. Our research focuses on developing a coarsening algorithm that uses CR to construct an adequate coarse variable set.

We consider testing the CR algorithm as outlined in the generalized AMG framework. We first attempted to implement the algorithm exactly as outlined by Falgout and Vassilevski. The results proved to be sensitive to parameters. After trying several different approaches we found a variant of the algorithm that is both robust and efficient in serial.

The results we obtain with this version are comparable with those of Oren Livne. The major drawback of this approach is that the smoother we use is multiplicative, which would be difficult to parallelize and thus not useful in practice, especially in the Hypre AMG package.

Our future work will focus on developing a version of CR that is robust and efficient and uses an additive fine point smoother. The primary issue to address is what to use as a “slowness” measure when choosing coarse points.

Discovery Center Display

Jedidiah Chow

Granada High School, Livermore, California

Mentor

Jean Shuler

ICCD

Summary

The goal of my project was to produce an interactive program that visitors to Lawrence Livermore National Laboratory could use to become informed about the computing goals and accomplishments of the Laboratory's Computation Directorate.

To simplify the project, I used only well-known programs. MicroSoft PowerPoint is the basis for the display program, and it proved adequate for navigating through various topics as well as produced aesthetically pleasing results. Basically, the program provides details about the history of supercomputing at Livermore, visual display equipment available at the Laboratory, and recent uses of the supercomputers there. The project was completed under the supervision of Terry Girill and Bob Howe.

General maintenance and modifications on the program will periodically be necessary (as with any program), and major updates will be accessed as time progresses and the information presented becomes out of date.

Multiresolution Contour Trees in Two Dimensions

Kree Cole-McLaughlin
University of Utah
Mentor
Valerio Pascucci
CASC

Summary

For the past three years, I have been working on algorithms for computing the contour tree and related data structures used in visualization of field data. Simply put, the contour tree is a representation of the connectivity of all of the level sets of a scalar field. More precisely, it is the graph obtained by contracting each connected component of every level set of a scalar field to a point. From the definition, one can see that the nodes of the graph correspond to critical points of the scalar field. Furthermore, it can be shown that if the domain of the field is simply connected, then the constructed graph has no cycles. If the domain has more complex topology, then resulting data structure is called the Reeb graph. Working with Valerio Pascucci of Livermore's Center for Advanced Scientific Computing and other collaborators, we have developed algorithms for efficiently computing the contour tree on domains of any dimension and the Reeb graph on two-dimensional domains.

Lawrence Livermore scientists are interested in analyzing very large datasets from physical simulation codes. These data can have an extremely large number of critical points. However, many of these critical points are often due to noise introduced by the simulation and are topologically less important. The topological noise in the data appears in the contour tree as many edges of the tree that span a very small range of field values. For the contour tree to be useful in the analysis of these datasets, we need some technique for isolating the important topology. In pursuit of this, I worked in the summer of 2003 on an algorithm for computing a multiresolution representation of the contour tree.

Although efficient techniques exist for computing the contour tree we have only been able to visualize a few nontrivial examples. Algorithms for the visualization of very large trees exist; however, they cannot be applied in our case. The problem with visualizing the contour tree is that we want to rank the nodes of the tree by their corresponding critical values. Imposing this condition makes it possible to construct examples of field that have contour trees with nonplanar embeddings. We have also shown that the multiresolution representation of the contour tree solves the problem of visualizing the contour tree.

We have developed an algorithm for computing the multiresolution contour tree for scalar fields in two-dimensional domains. I implemented the algorithm and produced several nontrivial examples using data from a two-dimensional combustion simulation. Using the multiresolution contour tree, we defined an embedding in three-dimensional space that is theoretically free of self intersections. The multiresolution contour tree allows for a visualization algorithm because it defines an order in which the edges of the tree can be drawn. Edges of coarser resolution are drawn, and then the edges of finer resolution are drawn until some threshold is reached.

Fast Implicit Computational Electromagnetics

Dylan Copeland
Texas A&M University
Mentor
Daniel White
CASC

Summary

An important problem for the EMSolve project is solving Maxwell's equations on very fine meshes arising from high-frequency integrated circuit applications. To produce simulations on a long time scale requires implicit time integration. The purpose of the summer assignment was to develop and test a multigrid code for this problem, with the hope that it would be more efficient than the simple conjugate gradient method with diagonal scaling.

In the beginning of the assignment, extensive tests were made with the multilevel code module (ML) software from Sandia National Laboratories. For sufficiently large problems, the Sandia software performed nearly twice as fast as the conjugate gradient method. However, the ML software uses algebraic multigrid, which does not take advantage of the rectilinear meshes associated with the integrated circuit problems in which we are interested.

Therefore, we developed a geometric multigrid code explicitly for rectilinear meshes, using hexahedral Nedelec edge elements. Solving Maxwell's equations implicitly poses some technical problems because of the large null space of the curl operator. To effectively eliminate these problems, we implemented a multigrid smoother developed by R. Hiptmair.

We ran numerical experiments with uniform and nonuniform mesh spacing, as well as multiple materials. In all cases, the geometric multigrid code performed much better than both the conjugate gradient method and explicit time stepping. For the simple case of a uniform mesh with only one material, our code was four times as fast as the conjugate gradient method. For more complex problems with two materials or nonuniform mesh spacing, the geometric multigrid code ran in about the same amount of time as in the uniform case, whereas the conjugate gradient method took more time than we could wait. Thus, our code is very efficient and performs well on practical problems similar to those posed by the integrated circuits we are interested in.

Although we have done some simple experiments with the geometric multigrid solver, it remains to be seen how well it performs on real problems. Ultimately, we would like to compare the performance of our code with that of the Sandia solver on a mesh arising from an integrated circuit. As an application of the code, we would like to apply multigrid as a preconditioner in the computation of eigenvalues for Maxwell's equations by subspace iteration.

Multigrid Preconditioning of Discontinuous Galerkin Methods for Elliptic Problems

Veselin Dobrev
Texas A&M University
Mentor
Panayot Vassilevski
CASC

Summary

The discontinuous Galerkin (DG) methods are relatively new tools for numerical solution of partial differential equations. Their main advantage compared to standard finite-element methods is that they are suitable for approximating solutions with discontinuities (jumps), shocks, or boundary layers. But they give rise to generally more complicated bilinear forms, larger discrete spaces, and linear systems with denser sparse matrices, which in some cases are nonsymmetric even for symmetric continuous problems. All of these factors increase the cost of solving the arising discrete problems, and thus efficient solution algorithms become very important. Multigrid methods are well known to be extremely efficient for preconditioning large linear systems obtained from discretizations of various types of PDEs. Therefore, they are a natural choice for the preconditioning of DG methods as well.

Three DG methods for elliptic problems were considered: the interior penalty (IP) method, the method of Baumann and Oden, and the nonsymmetric interior penalty Galerkin (NIPG) method. They were implemented for general tetrahedral meshes, and the code was parallelized using the Message Passing Interface (MPI) and the Hypre library.

Using the developed code, the convergence rates of the methods were tested numerically with piecewise linear and piecewise quadratic functions. The results for the latter agree with the theoretically known estimates, but for the former, we observe optimal convergence for all three methods, which has been proven only for the IP method.

The performance of two multigrid preconditioners was tested. Both of them use the natural embeddings of the discontinuous finite element spaces to define the interpolation/restriction, which leads to purely local operators. The first preconditioner tested was a variable V-cycle multigrid with diagonal smoother, whose diagonal entries are the row-sums of the absolute values of the elements of the matrix. The second one was V-cycle multigrid in which the ParaSails preconditioner from the Hypre library was used as the smoother. The first observation in these tests was that neither of the two preconditioners worked well for the method of Baumann and Oden, which requires further investigation. With the other two methods (IP and NIPG), both preconditioners performed very well, in that the number of iterations in the iterative method (GMRES for NIPG, PCG for IP) was bounded uniformly for all refinement levels.

Interesting directions for further studies are to understand better why the multigrid preconditioners failed to perform well for the method of Baumann and Oden and then to try to design better smoother and/or interpolation operators. Another interesting problem is to justify theoretically the numerically observed optimal convergence rates for piecewise linear functions. Last, but not least, the parallel scalability of the resulting algorithms should be tested and some possible improvements implemented.

This work was performed in collaboration with Raytcho Lazarov (Texas A&M University) and Panayot Vassilevski (Center for Advanced Scientific Computing, Lawrence Livermore National Laboratory).

Determining the Phenotypic Expression of Cytochrome P4501A2 Using Caffeine as a Probe

Tammi Duncan
Diné College
Mentor
Michael Malfatti
Lawrence Livermore National Laboratory

Summary

Cytochrome P4501A2 (CYP1A2) is involved in the metabolism of many compounds, including 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) and caffeine. PhIP is a heterocylic amine that is formed in certain cooked skeletal meats such as beef, chicken, pork, and fish when prepared under common cooking practices and may potentially become carcinogenic to humans when metabolized by CYP1A2. CYP1A2 activity can vary among individuals, which could alter the activation rate of compounds like PhIP. Therefore, to better evaluate the potential risks from PhIP exposure and bioactivation, determining individual expression of CYP1A2 is important.

Earlier studies have shown that caffeine is most commonly used as a probe. Therefore, we used CYP1A2 mediated caffeine metabolism to determine if individuals have a fast or slow CYP1A2 phenotype. We tested the rate of caffeine metabolism using saliva samples from volunteers who were dosed orally with 100 mg of caffeine. The samples were collected at different times (0 hr, 4 hr, 6 hr, 8 hr). First, the samples were collected and prepared for HPLC analysis. Second, the samples were injected into the HPLC to separate the caffeine metabolites. Caffeine and the CYP1A2 mediated caffeine metabolite, paraxanthine (17X), were identified and quantified based on retention time with known standards. Based on our data of paraxanthine to caffeine ratios from saliva samples collected at 6 hrs, generalizations could be made that designate individuals as fast, slow, or intermediate CYP1A2 metabolizers (see Figure 1). Out of 23 individuals tested, 8 were designated as fast, 10 were slow, and 5 were intermediate (see Table 1).

Overall the results of this study have furthered our understanding of carcinogens produced by our body. Future studies may indicate whether individuals with high CYP1A2 activity will be more susceptible to the carcinogenic risks associated with PhIP exposure.

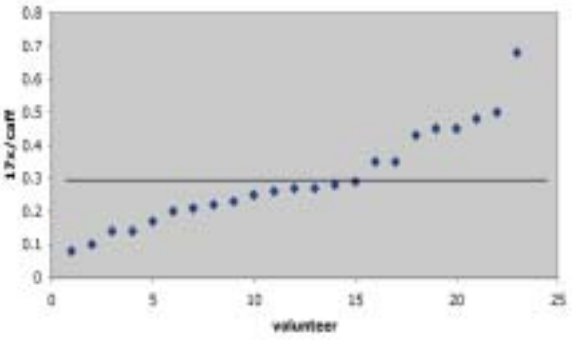


Fig. 1

Subject	17X/Caffeine	CYP1A2 Activity
1	-0.14	slow
2	0.08	slow
3	0.1	slow
4	0.14	slow
5	0.17	slow
6	0.2	slow
7	0.21	slow
8	0.22	slow
9	0.23	slow
10	0.25	slow
11	0.26	intermediate
12	0.27	intermediate
13	0.27	intermediate
14	0.28	intermediate
15	0.29	intermediate
16	0.35	fast
17	0.35	fast
18	0.43	fast
19	0.45	fast
20	0.45	fast
21	0.48	fast
22	0.5	fast
23	0.68	fast

Table 1

The Effect of Quantum Computation on Hash Functions

Emily Eder
Granada High School
Mentor
Terry Brugger
NAIC

Summary

Quantum computers will revolutionize computer science in the next 20 years. Conventional computer security procedures, currently suitable for classical computers, will become out of date. I show how the hash function, a central feature of message authentication, will be affected by quantum computers. Attacks on hash functions using classical computers are impractical because of space and time limitations. I developed a quantum algorithm that drastically reduces space and time requirements, both of which would make message authentication vulnerable to hackers with access to quantum computers. A central feature of the space reduction is the change from exponential to polynomial scaling.

The CH19 Form Designer

Abel Gezahegne
University of California, Davis
Mentor
Paul Amala
DCOM

ISCR Institute for Scientific Computing Research

Summary

The project was motivated by the need for biologists working on the chromosome 19 project to submit their findings either from experimental results or from other sources to a central database. A predefined interface they can use to submit results would either be too restrictive, too verbose, or even impossible in some cases because of the ad hoc nature of the experiments as well as the fact not all experiments can be predefined. Therefore, the objective of this project was to design and implement an application that the biologists may use to create their own interface tailored to fit any of the particular experiments they are performing. Using this custom interface, they can submit only the necessary information—nothing less, nothing more.

Under the direction of my supervisor, Tim Harsch, and two mentors, Amber Marsh and Julie Pitt, we designed and implemented an application in Java Swing that would allow the user to design a custom interface (form), use it to submit results to the database, as well as view previous submissions. To design a form, the user has a selection of predefined items that would be used to create an interface, for example, dropdown menus, multiselect lists, file selectors, text areas, and several others. The user will define the attribute of each of these items that make up the interface as well as the fields within the items. Once the interface is complete, the user has the option to save it as a template for future use. The user can then fill data in the designed form and submit the data, provided all the items have been filled with appropriate data. For each submission, the user is given a submission identification number that can be used for various purposes, including to view each submission.

As with any other application, modifications and improvements will be made to the CH19 Form Designer in the future. Aside from user-demanded improvements, some needs such as adding a date picker for creating a form, are already apparent for future versions. Functionality can also be improved so that users can not only view previous submissions but also update them.

Tools for a Statistical Package

Darryl Griffin-Simmons
Purdue University
Mentors
Dean Williams
&
Charles Doutriaux
EEBI

Summary

We seek to address the challenges of developing knowledge from global Earth system models by using statistics tools for analysis. Statistics tools are used in virtually every scientific project for analyzing and better understanding scientific data. The seamless integration of the Climate Data Analysis Tool (CDAT) and the Salstat statistical package provides climate scientists with a user-friendly tool for model diagnosis.

While at LLNL, I learned how to use CDAT software infrastructure subsystems, including:

- climate data management systems,
- a long masked numerical array operation called masked array, and
- Visualization and Control System.

I also modified the Salstat statistical package and worked with the CDAT graphical user interface (VCDAT).

Using SAMRAI and PETSc with the Immersed Boundary Method

Boyce Griffith
Courant Institute of Mathematical Sciences, New York University

Mentor

Richard Hornung
CASC

Summary

The immersed boundary (IB) method provides a mathematical and computational framework for addressing problems involving fluid–structure interaction and has proved to be especially useful in simulating biological fluid dynamics. To improve the efficiency of highly resolved IB computations, we have been developing an adaptive, parallel version of the IB method using the SAMRAI framework. Additionally, to overcome severe timestep restrictions found when using semiimplicit and explicit timestepping with the IB method, we are using PETSc to develop new implicit timestepping schemes. These methods should have the additional benefit of reducing splitting errors.

The work performed this summer is a continuation of work that began last summer as a Department of Energy Computational Science Graduate Fellowship laboratory practicum project.

The IB method describes the interaction of a viscous, incompressible fluid and an elastic structure. The approach used in the IB method avoids the necessity of remeshing the computational domain each timestep. It does so by describing the fluid on a Cartesian grid and the structure on a curvilinear mesh that moves with the fluid. A smoothed approximation to the Dirac delta function is used to connect these two computational domains. Using this smoothed delta function, quantities such as velocity may be interpolated from the Cartesian grid to the curvilinear mesh, and densities or distributions such as force may be spread from the curvilinear mesh to the Cartesian grid.

This summer, we have successfully implemented an adaptive version of the IB method using SAMRAI. Additionally, we have set up most of the data structures needed to interface this SAMRAI-based IB code with PETSc. We hope that we can take advantage of PETSc's facility for solving systems of nonlinear equations to implement fully implicit timestepping.

We intend to use this software with the 3D fluid-mechanical heart model of Peskin and McQueen. Additionally, we are interested in using the IB framework for an electrophysiological heart model with realistic (complex) geometry.

We are also interested in fully implicit timestepping. The PETSc interface developed this summer should facilitate this work.

Block-Matching for Video Object Tracking

Aglika Gyaourova
University of Nevada, Reno
Mentor
Chandrika Kamath
CASC

Summary

The growth of computer memory and processor speed has caused the expansion of automated systems for everyday applications. Computer vision systems constitute one example of such intensive computations systems, whose use becomes more feasible with the advances in computing power. Computer vision targets processing and analyzing visual information, which in general is characterized by large size and complex structure. The development of computer vision is hindered by the lack of complete understanding of how human beings process visual information. However, computer vision systems have already been incorporated in many practical applications (e.g., surveillance systems, medical imaging, robot navigation, and identity verification systems).

Object tracking is a key computer vision topic, which aims at detecting the position of a moving object from a video sequence. Models that describe road traffic patterns can be helpful in detection or prevention of uncommon and dangerous situations. Such models can be built by the use of motion detection algorithms applied on video data. Block-matching is a standard technique for encoding motion in video compression algorithms. It aims at detecting motion between two images viewed as a block. Each block of pixels in the current frame is matched to the best corresponding block of pixels in the destination frame.

We explored the capabilities of the block-matching algorithm applied to object tracking. The goal of our experiments is twofold: to explore the abilities of the block-matching algorithm on low-resolution and low-frame-rate video and to improve the motion detection performance by the use of different search techniques during the process of block matching. Our assumption for the process of video capturing is a motionless airborne camera. Although our experiments were made on tracking road vehicles (e.g., cars, trucks, motorcycles, and bicycles), the algorithm is general enough to be applied for tracking people or other moving objects.

Our experiments proved that the block-matching algorithm can be successfully used for object-tracking purposes from low-resolution and low-frame-rate video data. We observed that different searching methods have only a small effect on the final results and the additional computational load is not great. In addition, we proposed a technique based on frame history, which successfully overcame false motion caused by small camera movements.

Enhancements in VisIt

Akira Haddox
University of the Pacific
Mentor
Hank Childs
DCOM

Summary

VisIt is a parallel visualization and graphical analysis tool for scientific data used by many projects at Lawrence Livermore National Laboratory. VisIt is being actively developed, with new abilities and enhancements being requested and added. VisIt also has its share of reported bugs that need fixing. Readers need to be written for VisIt to support new users who use their own data formats. To meet other users' needs, various enhancements and readers must be coded, and bugs need to be fixed.

I implemented readers to support formats for Cosmos (LLNL's A Division), Mili (LLNL's B Division), and Boxlib (LLNL's A Division and Lawrence Berkeley National Laboratory). A new ThreeSlice operator was added that efficiently takes three orthogonal slices, aligned to the *x*, *y*, and *z* axes. I added another operator that removes cells from a dataset. A point tool was added to add interaction for the ThreeSlice operator, Streamline plot, and future point-based interactions. I contributed to speed enhancements in areas of mesh construction and contouring. I added several enhancements in the two-dimensional zooming interaction. Bugs were fixed in the KullLite reader, zooming interactors, mesh metrics, and the Box operator. To support the new Mili file reader, I implemented a parallel preprocessing tool to generate the metadata that VisIt needs, as well as added to VisIt an unstructured domain boundary exchange structure to create boundary information for datasets in serial or parallel.

VisIt is currently enhancing its AMR support, and when the support is completed, the Boxlib reader will be updated to reflect those changes to take full advantage of the AMR support. The unstructured domain boundary data structures will be expanded to exchange mixed materials and variables when a file format is added that requires that support.

Porting CALE to wxWindows

Jeffrey Hagelberg
Purdue University
Mentor
Paul Amala
DCOM

Summary

CALE (C Arbitrary Lagrangian–Eulerian) is a hydrodynamics simulation code written in C that was originally intended to run only on Unix and Linux systems. Over time, different versions have been created for various platforms such as the Macintosh. The primary goal of this project was to create a new version of CALE that would run on the Windows platform. Early in the project, it was decided that the best way to do this would be to use a platform-independent library so that we would be able to unify all of the various versions of CALE into one version that could run on all major platforms.

When we started, the first and most fundamental question we came upon was deciding what platform-independent package to use and what compiler to use. The compiler we chose was MinGW running under Windows 2000. MinGW is free and supports almost all of the standard UNIX header files. This universality was a huge advantage since it meant that most of the original CALE code did not have to be modified. The platform-independent package we decided to use was wxWindows. wxWindows was chosen because it is open source, robust, and provides support for drawing low-level graphics primitives such as lines and polygons in addition to providing support for creating dialogs, frames, and windows. Although it is primarily a graphics library, wxWindows also allows threads, files, and a variety of other things to be treated in a platform-independent manner. The new version of CALE was written in C++ because wxWindows was written in C++.

When implementing the wxWindows version of CALE, we had the command–prompt interface to CALE and the graphics window run in separate threads. This was challenging because the original version of CALE was single-threaded. We also wanted to keep the original CALE code and the new graphics code as separate as possible to minimize the coupling between the two. We achieved this separation by creating one class for each drawing event that occurred. For example, a class called FontSizeChanger was created that knows how to change the font size. All of these classes are subclasses of a base class called GraphicsEvent. All of the GraphicsEvent classes do not correspond directly to a low-level drawing call. In fact, the drawing they do could be arbitrarily complex. This distinction highlights one of the big advantages of this approach. It hides the details about how the drawing is actually accomplished and thus simplifies the task of implementing the drawing while making the drawing process less error-prone. An additional goal in creating the new version was to modernize the user interface. This came automatically when all of the dialogs were rewritten.

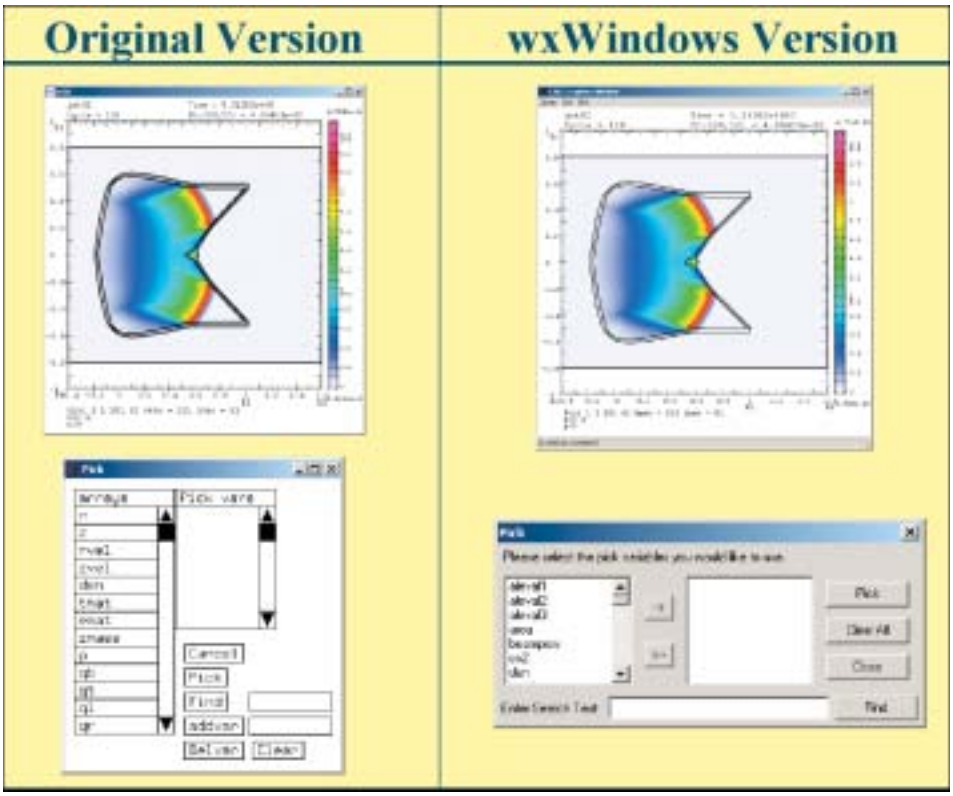
As of now, the wxWindows version of CALE only runs on Windows. Work is currently under way to make this version run on the Macintosh and Linux platforms as well. Additionally, more work needs to be done with regard to memory management and thread synchronization.

Continued

Jeffrey Hagelberg
Purdue University

Summary continued

The diagram below shows a comparison between some of the graphics features in the old version of CALE and those in the version using wxWindows. This example shows a typical CALE problem run.



Comparison of original version and wxWindows version of CALE

BlueGene/L: Providing Large Scalability for Scientific Computing

Amy Henning
University of California, Santa Cruz
Mentor
Don Dossa
CASC

Summary

BlueGene/L (BG/L) is a custom supercomputer providing massively parallel scalability across 65,536 compute nodes and 1024 I/O nodes with a target peak performance of 360 teraflop/s. Each compute node is composed of two system-on-a-chip PowerPC 440 processors with a nominal clock frequency of 700 MHz and an associated double pipe floating point unit, along with 512 MB of local memory. All compute nodes are interconnected by a three-dimensional torus, a global tree, and a barrier and interrupt network.

BlueGene/L is designed to deliver large-scale scientific computing to allow direct comparisons with real scientific experimentation and will result in faster runs than on previous supercomputing platforms. It is our goal to first run science applications and solvers and ultimately to incorporate techniques learned for scaling applications to 64,000 nodes to stockpile-stewardship codes. Efforts involved in developing BG/L will also present a promising architecture to be further scaled to the petaflop/s level.

For early software development purposes, IBM implemented a simulation environment for remote accessibility. The simulator (BGLsim) is a 64-way cluster running on a Linux microkernel. Efforts on both the simulator and real BG/L hardware involve porting, building, and evaluating scientific applications such as an optimized and "vanilla" version of the hydrodynamics code sPPM and a shock-wave solid-mechanics code, CTH, developed by Sandia National Laboratories.

Before we could begin porting to the simulator, it was necessary to implement a running version on a more stable and local platform such as the IBM SP and the Livermore Center for Advanced Scientific Computing's (CASC's) Linux cluster. BG/L uses the latest IBM XL compilers and the MPICH package, which, in turn, require us to use and build those packages on select platforms. For sPPM, we generated successful results on the CASC Linux cluster using gnu compilers and MPICH2 and on a Power3 SP (ASCI Frost) with XL compilers and MPICH2. On a smaller PowerPC cluster (ASCI Smurf), we were able to run a version of CTH. These applications could then be ported to BGLsim. It was essential to monitor the memory usage of the applications since BGLsim has a limitation of 256 MB per simulated node. Some modifications to source codes and compilation options were necessary for a temporary work-around. Working on an experimental simulation environment requires us to meticulously detect bugs associated with such features as the Fortran runtime and Makefiles hierarchies. Building large scientific applications on a variety of platforms presents a unique set of challenges and allows for systematic approaches to tackling bugs during compilation and runtime over time.

Work on the simulator resulted in successful runs of the vanilla version of sPPM with single precision using MPI and a single task run. We have a program provided by Bruce Curtis that verifies the correctness of output and has been thoroughly tested on Frost and Berg. It is our intent to use this checker when we obtain results from a double

Continued

Amy Henning
University of California, Santa Cruz

Summary continued

precision run of sPPM on BG/L. We are still in the development phases for CTH on BGLsim. Issues with the XL Fortran runtime and a few library routines need to be resolved by IBM before further progress.

Once we had a working version of an application running on the simulator, such as sPPM, we promptly ported it to actual BG/L hardware at IBM Watson Research Center in Yorktown Heights, New York. By running the application on the hardware, we can further evaluate its performance by obtaining timing measurements. Initial runs of an optimized version of sPPM show a 54% improvement in speed by simply using the XL compilers over GNU. Having employed only one hummer FPU at this stage of testing, we anticipate definite improvement in all applications with the utilization of double hummer FPU technology. With further runs to scale on the hardware, we will be able to formulate comparisons with other platforms and demonstrate the advantages of this architecture.

Continuous interests in BG/L led me to participate in research on job scheduler performance with Andy Yoo from CASC. Because of the massive number of compute nodes, having an efficient scheduler is essential. We are evaluating First-Come, First-Serve and Conservative-Backfilling algorithms in a small simulation environment evolved by using a bimodal hyper exponential distribution function as the basis for calculated results. The simulator generates performance measurements of the schedulers, including system use and response time. In order to conduct varied test cases, we made modifications to the simulator to allow for more parameter specifications. Such options include setting the degree of distribution functions or categorizing jobs by size or execution duration. We also added features to generate data files used for further visualization of our results. It is our prediction that the backfilling algorithm will provide minimal effects to system use and will not be efficient enough to implement on BG/L.

The ASCI BlueGene/L team has allowed me an opportunity to participate in countless meetings, telephone and video conferences, and visits to IBM Watson Research Center. Since BG/L is a collaborative effort between Lawrence Livermore and IBM, I attended a BG/L workshop in March where we were given a tutorial of the BGLsim and presented with information of progress by IBM on BG/L. Being part of the BG/L team, I attended the Conference on High-Speed Computing, where I had the opportunity to become acquainted with areas of research that ranged from memory bandwidth improvements to new hardware approaches to supercomputing. This conference gave me a broader prospective of what types of research I could possibly be involved in at the graduate level. At the conference, I was also able to present my work on BG/L by delivering a poster presentation and sharing what I've been involved in at LLNL.

In July, we had our annual review of the project. The review allowed a chance to interact with other collaborators and to get a more in-depth view of the overall efforts involved

Continued

Amy Henning
University of California, Santa Cruz

Summary continued

with BG/L. These past 6 months has given me the privilege to witness the intricate challenges and overwhelming rewards of creating a state-of-the-art supercomputer.

With Andy Yoo I am writing a paper entitled, “The Effectiveness of Backfilling on Improving Utilization of BlueGene/L,” anticipated in 2004.

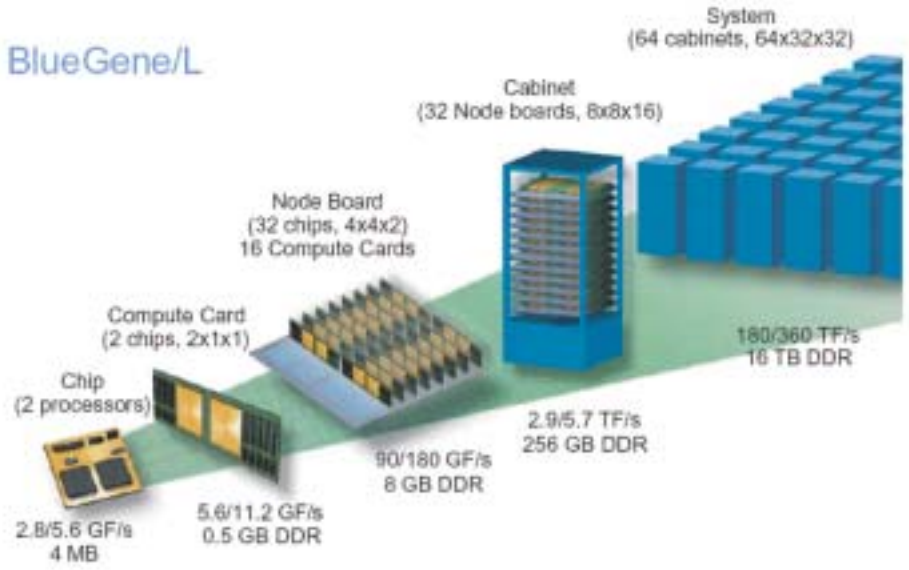


Fig. 1. The diagram of the BlueGene/L layout demonstrates the progressive development stages of the BlueGene/L hardware from a single application-specific integrated circuit chip to a full system.

Parallel Image Compositing Application Program Interface (PICA) Support for Chromium

Mike Houston
Stanford University
Mentor
Randy Frank
ICCD

Summary

The Parallel Image Compositing API (PICA) is designed to allow large parallel rendering applications to use a standard application program interface (API) for delivering content to various display sources using different imaging compositing systems. Chromium is a stream-processing framework for interactive rendering on clusters developed at Stanford University as part of a Department of Energy VIEWS grant. The objective was to add base PICA support to Chromium as well as provide an initial software compositing layer. Besides testing the feasibility of the current revision of PICA, this project sought to provide a compositing layer for use with VisIT and other LLNL visualization packages.

PICA support is now in the main Chromium CVS branch. PICA support is implemented as Chromium extensions (crPica), much as the parallel programming extensions such as semaphores and barriers are implemented. Several of the stream-processing units (SPUs) in Chromium required modification to correctly support the new API. Most just require dummy functions to be inserted to report errors, e.g., the crPICA calls really do not make sense in the “render” SPU. The “tilesort” SPU needed to be modified to correctly scale and position pixel data for large displays as well as correctly report the display size. The current, non-PICA compositing SPUs like “binaryswap” and “readback” can now correctly render to large tiled displays.

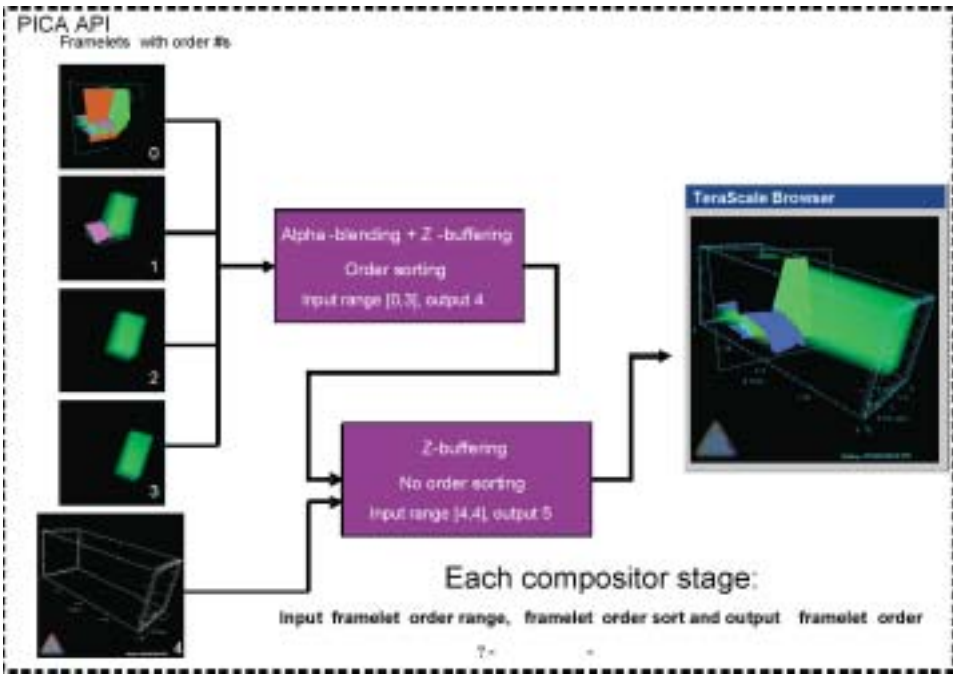
While implementing a software composting system for Chromium and PICA, a “composite” SPU, it became clear that the Chromium's networking layer had several major bugs. Most of the bugs where exposed because of the need to heavily use Chromium's out-of-band layer. High-speed network layers such as Myrinet and Quadrics were incorrectly using buffer pools, which were used to avoid expensive memory reallocation in DMA space. These bugs have been fixed as part of this effort, but several still remain, most notably the lack of support for nonblocking “connects” and “sends” in the most heavily used TCP/IP layer. Nonblocking communication is required for the compositing nodes to handle ordering problems during alpha compositing and to use more efficient compositing algorithms like “binaryswap” and SLIC.

I expect to finish fixing and optimizing Chromium's network layer by the end of the year. Besides generally improving the performance of Chromium, this work will allow me to finish a software composite layer as well as allow the design of SPUs that require nonblocking communication. Once hardware compositing systems designed for PICA become available from vendors, it will be interesting to see how they fit into Chromium and its PICA layer.

Continued

Mike Houston
Stanford University

Summary continued



A simple “composite” example using the TeraScale browser. Several nodes are handling the volume rendering, and one is handling the bounding box. The composite requires two stages, one stage for alpha blending the volume rendering and one stage for depth compositing the bounding box.

Volumetric Data Analysis and Compression

Lorenzo Ibarria
Georgia Tech
Mentor
Terence Critchlow
CASC

Summary

Scientific simulations produce large amounts of data, on the order of many gigabytes. With datasets this large, the storage space is costly, so compression must be addressed. Conventional compression techniques do not take advantage of the knowledge of the data or of its multidimensional coherence, and several compression methods that could use this knowledge are not ready for such a large amount of data. We have chosen to transform the data with a predictor and to apply a symbolic compression to the corrections. There are several kinds of predictors: interpolating, extrapolating, and fitting predictors. Our objective this summer has been to evaluate the effectiveness of each kind of predictor in the data, looking at factors such as the dimension of the predictor, its progressiveness, and how it performed against quantization.

Our tests were performed in the four-dimensional PPM simulation that represents two fluids mixing after a Mach 1 shock and in a series of two-dimensional images called the Miranda data, where two fluids of different density interact. Our results are not conclusive, and we advise to use a predictor of the highest order possible within the data features. The more quantized the data, using a simpler—and possibly an extrapolating—predictor is better.

Our last test was performed on a local n-dimensional extrapolating predictor, because it is useful in out-of-core compression and decompression. We studied its correction patterns and found that it had a much lower error bound along the axis. The ability to guess the performance of a predictor is a handy feature, because that information combined with a context arithmetic encoder can lead to great gains. Our naïve approach at using the prediction as a context resulted in a reported 25% increase in compression.

In the future, we plan to apply these ideas not only to large dataset compression and decompression, but also to their form of information, such as isosurfaces. We have seen that our predictors work well even outside a regular grid, and we are eager to follow leads into the isosurface time-dependent compression.

For more information on our predictor, see L. Ibarria, L. Lindstrom, X. Rossignac, and X. Szymczak, "Out-of-core compression of large n-dimensional scalar fields," *Proceedings of Eurographics 2003*.

Streaming Meshes

Martin Isenburg
University of North Carolina
Mentor
Peter Lindstrom
CASC

Summary

The standard indexed format for polygonal meshes is an array of *n* vertex records (e.g., *xyz*-coordinates, pressure values, surface normals, etc.) that is followed by *m* triangle records that index into the vertex array. When working with large meshes, this representation hampers efficient processing of the mesh data. In our work, we propose a novel mesh representation we call streaming meshes. In this representation, the vertex and triangle records appear interleaved, such that vertex records appear nearby the triangle records that use them. In addition, streaming meshes provide information about when vertices are no longer used (when a vertex is "finalized"). Large meshes in this representation can be efficiently streamed through limited memory, and seamless connectivity information can be made available along the active elements (those vertices—and their surrounding triangles—that have been introduced but have not yet been finalized).

In our work, we define streaming meshes, characterize their properties such as stream width (the maximal number of elements that are active at the same time) or the stream age (the longest time span that a particular element remains active). We show that certain processing methods already greatly benefit from a low stream width, while others also require the stream age to be low. In addition, we figured out how to create good streaming meshes from preexisting standard indexed meshes using an out-of-core method (a method that operates using only a limited amount of main memory, with the bulk of the data residing on disk). We also implemented an SM reader/SM writer application program interface (API) that allows mesh generation algorithms such as Marching Cubes to output a mesh directly in a streaming format. Finally, we designed a compression scheme that can compress streaming meshes on the fly in the order written through the API. The main challenge here, compared to previous work in mesh compression, was to account for the fact that the order in which the mesh is traversed is now determined by the process that writes the streaming mesh. Previous work in mesh compression always assumed that the entire mesh was available so that the compressor could choose its own traversal order, usually such that compression is maximal.

We have now a fully functional API for reading and writing streaming meshes. In addition, we have a mechanism that converts these streaming meshes on the fly into a processing sequence. In recent joint work with Stefan Gumhold and Jack Snoeyink, we have shown that processing sequences are useful for efficient out-of-core processing of gigantic mesh data sets. Now that we can efficiently generate processing sequences directly from the output of, for example, a Marching Cubes algorithm, we can streamline the entire mesh-processing pipeline. This ability will significantly speed up a previously cumbersome workflow. With streaming processing, we can likely rely solely on CPUs and no longer be bound to I/Os.

Babel: Now Serving Java

Sarah Knoop
University of Wisconsin, Madison
Mentor
Gary Kumfert
CASC

Summary

Our aim this summer was to enable Java on the server side Java of the language interoperability tool Babel. Previously, Babel facilitated communication among C, C++, Fortran 77, Fortran 90, and Python on both the client and server sides but the tool operated only on the client side of Java. With the work completed this summer, Babel is much closer to presenting a full set of services.

Babel connects to Java through the Java Native Interface (JNI), a Java platform feature that allows a Java code to interact/interoperate with the host environment of the machine (i.e., the operating system, native libraries, etc.). JNI is a two-way interface that Java programs to invoke native code (usually C or C++) and vice versa. For the complete server side Java functionality, the server-side bindings need to accomplish successful and type-safe conversions between the incoming Babel IOR (C-style) types and Java types. The converted Java types are the arguments that get passed to the Java server implementation. The reverse must be done with the return value from the Java code. It must be converted back to an appropriate Babel IOR type.

All Java server tests involving the above data types are building and passing successfully. We also have a good start on handling Object arguments; however, an unexpected error with the client side prevented full testing.

The following argument and return types have yet to be fully enabled and tested: Objects, Arrays, Enums, Opaques, Classes, and Interfaces. Also, a more succinct mechanism for server-side exception handling should be established and tested. Additionally, the management of local references in the server-wrapper code should be more carefully implemented. The generation code is designed so that much of the basic functionality needed by these additional tasks is already in place. Calling basic functions and implementing the details with respect to each task should be relatively easy to insert in the appropriate places, once they have been conceptually thought through.

Dual Least-Squares Methods for Computational Electromagnetics

Tzanio Kolev
Texas A&M University
Mentor
Panayot Vassilevski
CASC

Summary

The goal of this project was to develop efficient algorithms for the numerical solution of problems arising from various electromagnetic models. Specifically, we are interested in the system of Maxwell's equations and the related eigenvalue problem. These are important in practical applications such as the computation of the electromagnetic field generated by prescribed current and charges or in the computation of the eigenmodes that will propagate through a given medium. The work was carried out in collaboration with Joseph Pasciak from Texas A&M University and Panayot Vassilevski and Daniel White of CASC.

Our approach is based on very weak variational formulations of the div-curl systems corresponding to the electrostatic and magnetostatic problems. The finite dimensional approximation is a negative-norm finite-element least-squares algorithm that uses different solution and test spaces. This algorithm allows for approximation of problems with low regularity, where the solution is only in L^2 and the data resides in various dual spaces. The solution operators for the above problems are further used to obtain an approximation to the eigenvalue problem.

The resulting discretization method has the advantages of avoiding potentials and the use of Nedelec spaces. In fact, we allow for the mixing of continuous and discontinuous approximation spaces of varying polynomial degrees. Additional advantages are that the matrix of the discrete system is uniformly equivalent to the mass matrix and that spurious eigenmodes are completely avoided. Finally, the dual inner products can be efficiently implemented using preconditioner for standard second-order problems (for example, a sweep of multigrid).

A computer program was developed that applies the method to the Maxwell equations and the eigenvalue problem in the frequency domain. It is written in C++ in the framework of the AggieFEM finite element library, which supports complex geometries, local refinement, multigrid preconditioning, and OpenGL visualization. The code is based on the solvers for the magnetostatic and electrostatic problems. It works on triangular, tetrahedral, and hexahedral meshes. It provides an eigenvalue solver, which allows for computation of blocks of eigenvalues, and a solver for the full-time harmonic system. A parallel version of this code was implemented based on MPI and CASC's Hypre preconditioning library. It uses matrices built in parcsr parallel format, Lobpcg as a parallel eigensolver, and BoomerAMG or ParaSails as a preconditioner.

Various tests were performed to confirm the theoretical estimates for the least-squares method. Numerical experiments for the eigenvalue problem on a linear accelerator induction cell were compared with the those from the EMSolve project in CASC, which is based on Nedelec elements. The general observation is that the efficiency of the method on hierarchically refined meshes (where geometric multigrid can be used) can be preserved on unstructured problems if one uses algebraic multigrid.

Creating a Graphical User Interface for a Herbivore Population Simulation

Magdalena Kowalska
Warsaw University

Mentors

Tanya Vassilevska
&
Tina Carlsen
CASC

Summary

As part of a Department of Energy project to assess risk to herbivore populations from petroleum exploration, a simulation of a herbivore population was created for researching the effect of habitat reduction and fragmentation. The goal of the project is to develop an ecological framework for evaluating effects of environmental pollution on population persistence. The model for the simulation incorporates spatial movement and social behavior together with a large amount of quantitative information. Incorporating all of the data in the model is possible only by using a discrete individual-based model. The model is a nonlinear discrete time map, which transforms a certain number of matrices at time n into the same number of matrices in time $n+1$. Each individual animal is represented in this model as a vector of the characteristic attributes of its species. The population is a matrix consisting of such vectors. The habitat of the population is divided into square cells; consequently, the environment is represented in the model as a matrix of vectors of attributes of one cell. The discrete maps contain both deterministic and probabilistic rules.

The program that is based on this model has a complex input. It consists of several files containing environmental and species data used to generate the initial population of animals and the initial state of every cell. To evaluate the outcome of a discrete simulation containing probabilistic rules, one conducts a large number of simulation experiments that require frequent manipulation of the input files and the output format. To conduct these simulation experiments, the user must have a detailed knowledge of the code and the nature of its parameters. The code's intended users, most of whom are ecologists, do not have this detailed knowledge and thus find it difficult or impossible to work with the code. The solution to this problem lies in the construction of a graphical user interface (GUI).

We decided that a front-end architecture is most suitable for the project because it prevents interference with the structure of the simulation program. A GUI created in this way is a “wrapper” for the simulation. It allows for fast and convenient generation of input. The GUI incorporates the data from all input files used by the simulation, facilitates their adjustment, and passes them to the simulation in the form of a configuration file. We created a GUI that is appropriate for different simulations based on similar format of the input data. Thus the GUI has some features that allow for editing the format of data. In addition, users have an option to choose which simulation they want to run. The tool not only allows for the rapid creation of an application from scratch but also provides a number of templates to start with.

We plan a distribution of the software.

Investigation of the Oblique Shock Richtmyer–Meshkov Instability

Marco Latini
California Institute of Technology
Mentor
Oleg Schilling
Lawrence Livermore National Laboratory

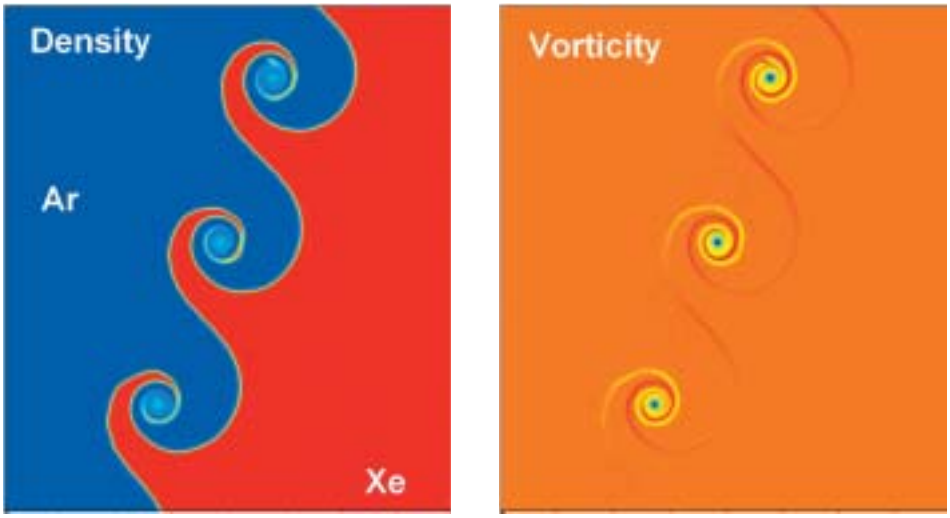
Summary

In the classical setup, the Richtmyer–Meshkov instability (RMI) is studied in the case of a normal shock interacting with a perturbed material interface. The shock deposits baroclinically generated vorticity on the interface, which drives the development of the instability and the formation of a mixing layer at the interface.

In the present work, we revisit the instability observed when an oblique shock interacts with a perturbed interface. We call this case the oblique shock Richtmyer–Meshkov instability. In our investigation, we use the weighted essentially nonoscillatory (WENO) high-order shock-capturing method and analysis to elucidate the complexity of the hydrodynamics resulting from the interplay of the Richtmyer–Meshkov and Kelvin–Helmholtz instabilities.

This study is motivated by current research and experiments on inertial confinement fusion (ICF) capsules. In an ICF capsule implosion, a laser drives a shock that compresses a deuterium–tritium (DT) mixture causing ignition and fusion. The mixing resulting from RMI and other instabilities limits the compression and the yield in energy. Oblique shocks may be caused by asymmetric drives. It is therefore important to study the hydrodynamics driving the oblique shock RMI.

In our study, numerical simulations were conducted using the WENO shock-capturing method developed at Brown University by Wai-Sun Don. This method provides unprecedented resolution of small-scale features through high-order reconstruction procedures. We present for the first time numerical simulations of the oblique shock



This figure represents density and vorticity plots for the oblique shock Richtmyer–Meshkov instability at time t = 2 ms after the passage of a Ma = 1.2 shock. The interface separates argon and xenon and was initially at an inclination of 75 degrees. The initial sinusoidal perturbation had an amplitude of a = 0.2 cm and k = 1.8888. The vorticity deposited by the shock drives the initial growth of the perturbation and the Kelvin–Helmholtz induced shear contributes to the formation of the roll ups. Numerical simulations are obtained using the WENO 5 method developed at Brown University.

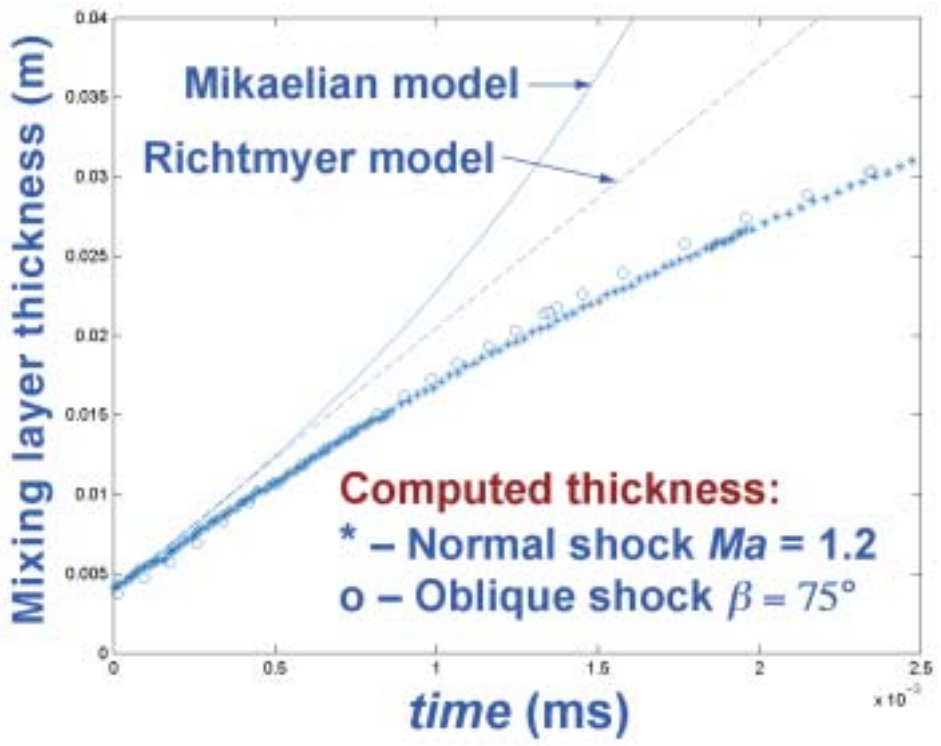
Continued

Marco Latini
California Institute of Technology

Summary continued

RMI, including a comparison to Mikaelian's prediction of the growth rate of the mixing layer. Our study suggests that Mikaelian's model captures the growth rate of the instability under limited circumstances, and we are currently working on improving the model. We identify the mechanisms driving the oblique shock RMI in the vorticity deposited baroclinically by the shock and the velocity shear as the shock passes. Furthermore, our results indicate that high-order reconstruction-evolution techniques provide conservative shock-capturing capability, as well as accurate resolution of small-scale flow features.

This investigation is one component of a larger program investigating physical and numerical aspects of complex hydrodynamics and turbulent mixing generated by accelerations.



In this figure we compare the growth rate of the mixing layer following refraction of a normal and oblique shock. In both instances the shock refracted at a speed of Ma = 1.2. We compare the computed growth rate of the normal shock case against the Richtmyer model and the oblique shock against the Mikaelian model. Our results suggest limited agreement with the Mikaelian model.

RETURN TO TOC

Creating a Standard, Extensible Mechanism for Native MPI Communication in Python MPI

Taylor Leese
University of San Francisco
Mentor
Pat Miller
CASC

Summary

Python MPI (pyMPI) is a Python interface to MPI and is designed to provide parallel operations for Python on distributed, parallel machines. There are certain difficulties in communicating data in pyMPI that are not present in C or Fortran. Python is an interpreted programming language while C and Fortran are compiled. Consequently, the data types of variables in Python are only available at execution time while in C and Fortran they are available at compile time. Since data types in Python are dynamically typed, it is necessary to compute these types during execution in order to ascertain the proper way of communicating them using MPI. Object serialization provides a solution for communicating dynamically typed variables since it isn't necessary to know the type of an object during serialization. However, the previous pyMPI design of using object serialization to communicate data in pyMPI creates unnecessary overhead. Removing the need to serialize all data in pyMPI was a major motivation in creating a new design for communicating basic Python types and Numeric Python (NumPY) arrays. Aside from the extra processing needed to serialize and de-serialize a Python object, the message sizes for serialized objects are much larger than the size of their respective Python data types. In turn, eliminating the need to use object serialization would improve the scalability of pyMPI and result in a design that creates a major improvement in overall performance.

To communicate data natively (without using serialization) using pyMPI, specific information about a particular Python data type must be known. Thus, a new design was created for pyMPI using a hash table (represented as a Python dictionary). Python objects are hashed using their Python type (PyTypeObject). Each entry in the hash table is a Python tuple containing three callable Python objects. Each callable Python object returns a Python tuple containing the information needed to either pack, unpack, or reduce the particular data type respectively. If a data type is not in the registry (the hash table represented as a Python dictionary), then the default is to communicate the data as a PyBaseObject_Type. Objects having PyBaseObject_Type or data types not in the registry default to using the cPickle module to communicate the data. Thus, the need for object serialization using cPickle has not been entirely removed from pyMPI, but instead, it is only used when absolutely necessary to communicate the data.

Initially, there are either four or five entries in the registry depending upon whether or not the NumPy package was installed on the system prior to compiling pyMPI. These types are PyFloat_Type, PyInt_Type, PyBaseObject_Type, PyString_Type, and possibly PyArray_Type if NumPy is installed. However, the system is not limited to these data types. Using a Python dictionary as a hash table allows the registry to be easily extended for new Python data types. Thus, the new design has created a standard and extensible mechanism for communicating Python data types without object serialization. It is straightforward for users to add their own Python data types to the

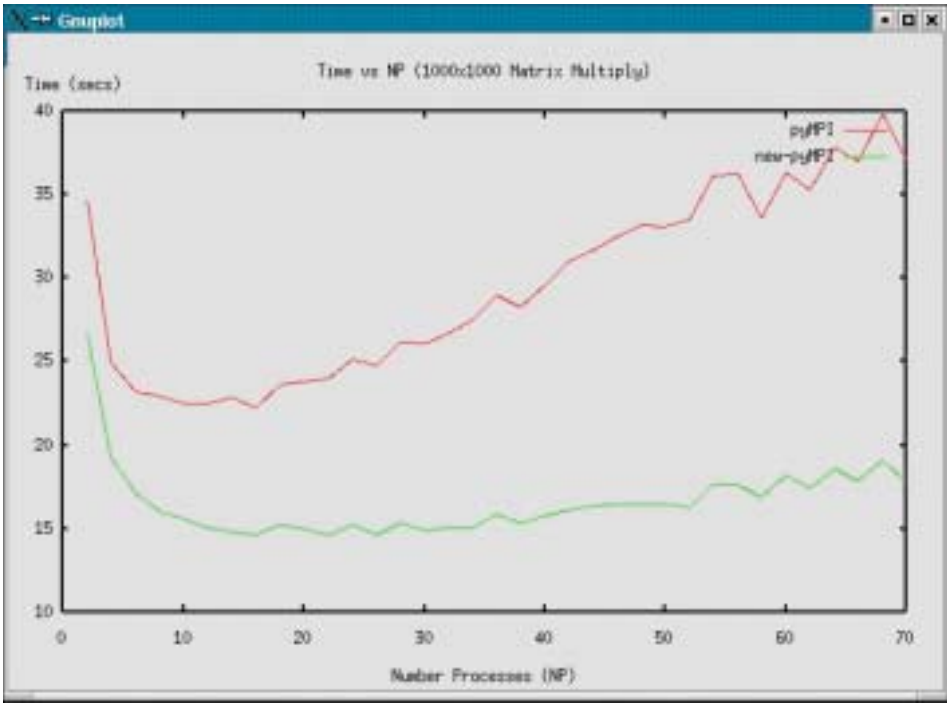
Continued

Taylor Leese
University of San Francisco

Summary continued

registry. The user must simply insert a tuple containing the three callable Python objects used to pack, unpack, and reduce the datatype into the pyMPI registry using the type of this new data type as the key. Currently, send, receive, sendrecv, broadcast, isend, and irecv all communicate data natively using this design.

Currently, there is no support for obtaining native communication in pyMPI for gather or scatter. This is an issue inherent to Python since the data being gathered or scattered is not necessarily composed of homogeneous types. Determining how to communicate nonhomogeneous data natively is an issue to be resolved at a later time. In the current version of pyMPI, this problem is handled by using object serialization since Python objects with nonhomogeneous data pose no problem for object serialization.



Time vs. number of processes for a 1000 × 1000 parallel matrix multiply. This graph shows the improvements in speed and scalability achieved by the new pyMPI design.

A Web-Based Genome Analysis Tool

Brian Lum
UC Berkeley
Mentor
Tim Harsch
EEBI

Summary

My main project was to design a web-based cross-genome analysis tool. Biologists use a tool that performs cross-genome analysis. Unfortunately, that program is built for a Unix system and is fairly complicated to use. I built a front end for them so they would have a user-friendly graphical interface.

Before I started on this project, I was given two other small projects that were mainly educational. These taught me about Perl DBI and CGI so that I could access databases and build a web interface. For my first project, I transformed a customer input into an XML format that was required by another analysis tool. The second project was editing a preexisting web interface to show more information from a database.

For my main project, I used the knowledge from the first two projects. I spent a great deal of time predicting all the ways user could give bad inputs and provided ways of handling those errors appropriately. The genome analysis tool takes a long time to run, so we want to check the input from users as much as possible beforehand, so we can catch the errors before using the program. After finishing that project, I made another analysis tool that performed the same operation, but for a different group.

Currently, the web-based interface does not give all the options of the tools. After talking to many biologists, we decided not to give those options for simplicity purposes. In the future, this might change. Also, there is currently a limited scope for which genomes are compared. We may decide to add different genomes.

Time-Varying Reeb Graphs

Ajith Mascarenhas
University of North Carolina, Chapel Hill
Mentor
Daniel Laney
CASC

Summary

Physical processes that are measured over time, or that are modeled and simulated on a computer, can produce large amounts of time-varying data that must be interpreted with the assistance of computational tools. A popular method is computing and visualizing isosurfaces of the data. Topological properties of the isosurface, such as the number of connected components, number of handles (genus), and number of voids, can be important aids in interpreting data. The Reeb graph is a useful topological structure to study these properties. Intuitively, the Reeb graph encodes topological changes to the isosurface as the isovalue is varied. In this project, we study the evolution of the Reeb graph of a time-varying function defined on S^3 × R and develop an algorithm to compute the time-varying Reeb graph for functions defined by piecewise linear interpolation from sampled data.

Research effort was largely directed toward development of the theory and algorithm to compute time-varying Reeb graphs. The result is a manuscript to be published at a future date. Implementation is in progress. This is the first work to address the study and computation of time-varying Reeb graphs. This research was in collaboration with Valerio Pascucci of Livermore's Center for Advanced Scientific Computing, Herbert Edelsbrunner and John Harer of Duke University, and Jack Snoeyink of the University of North Carolina at Chapel Hill.

Implementation and application to large time-varying data sets is in progress. The work will be carried out at the University of North Carolina.

Autonomous Motion Segmentation of Multiple Objects in Low-Resolution Video Using Variational Level Sets

Mark Moelich
University of California, Los Angeles
Mentor
Chandrika Kamath
CASC

Summary

My doctoral research at UCLA has focused on variational and statistical methods for tracking objects in video sequences. The group I work with at LLNL is primarily involved with data mining but has recently become interested in extracting and mining information from long, low-resolution (in both space and time) video sequences. The primary application is surveillance, and the emphasis on low-resolution was motivated by Livermore's experience handling large data sets. My objective for the summer was to develop an algorithm that is capable of isolating (segmenting) moving objects in low-resolution video sequences.

My approach was to use a region-based, variational level set method to isolate the moving objects. The algorithm that I developed is based on a functional (or energy) that depends upon the content of the current and preceding two frames of video. The functional is minimized to provide the desired segmentation by evolving a PDE in the level set framework. Although some loss in performance is expected at low resolution, this approach works well for the intended class of video sequences. In contrast to traditional image-processing techniques, our approach has a strong mathematical basis, and unlike local methods such as optical flow estimation, it is effective at low frame rates. This method also avoids some of the problems involved with using background subtraction.

There are several ways to extend the algorithm. The most important are to develop an algorithm that provides a frame-to-frame correlation of the objects and to develop the data-mining algorithms for which this algorithm was intended.

Adaptive Algebraic Multigrid and Frequency Filters

Arne Naegel
University of Heidelberg
Mentor
Rob Falgout
CASC

Summary

The need for large-scale simulation and the power of current and future parallel architectures make the development of state-of-the-art solution techniques an important task for today's research. A key question arising in this context is the solution of large linear systems that arise when problems formulated in terms of (systems of) partial differential equations are discretized.

By the algebraic nature of the problem, this question is closely linked to the spectrum of the matrices. The class of multigrid or multilevel methods incorporates the idea of treating all frequencies of this spectrum equally and therefore has optimal or near-optimal efficiency and convergence properties. For simple model problems with underlying geometric information, the original geometric multigrid has proved to be highly effective. Because these methods are somewhat limited when applied to large unstructured problems, algebraic multigrid (AMG) was developed, generalizing the basic idea, but without the need to rely on any geometric information. Briefly described, the key idea of the method we investigated and of this approach in general is to decompose the space into two parts. The first subspace is one in which a basic solver, the so-called smoother, can effectively be applied; it can typically be described by eigenvectors associated with large eigenvalues. Its complementary space can consequently consist of eigenvectors associated with small eigenvalues, and we construct solvers for this space using one or several of these algebraically smooth vectors as representatives or test vectors to construct an efficient method.

The UG software toolbox was used to implement an adaptive AMG version that is able to identify the space where the smoother is not efficient, to generate an appropriate test vector, and finally to improve itself using this information. Its basic principles were derived from a work by Ch. Wagner, whose method has been generalized to be applied to poorly scaled matrices. A variety of numerical tests for representative problems were performed to compare the original method and its modification. In the cases where the original method fails, convergence was established or convergence factors were improved by one order of magnitude. We additionally examined the behavior of the new method in possible frameworks for self-correcting schemes.

Though already responsible for some modest success, the method promises an even bigger potential and should thus be considered a work in progress. Through our analysis, we were able to identify strengths and weaknesses of the adaptive AMG approach taken and gained valuable insight into the crucial role of the test vector. Research will be continued in Heidelberg, and future work will be done in continuing collaboration with Robert Falgout and the Scalable Linear Solvers Group at Livermore.

3D Morse-Smale Complexes

Vijay Natarajan
Duke University
Mentor
Daniel Laney
CASC

Summary

Morse functions are used in differential topology to study the topology of manifolds. We use the results from Morse theory to study the topological features of natural phenomena. For example, in x-ray crystallography, we can reconstruct the geometry by following ridges connecting electron density maxima. A Morse–Smale complex partitions the 3D space into cells with similar gradient flow characteristics. In earlier work, we gave a combinatorial algorithm to construct the Morse–Smale complex for piecewise linear domains. In this project, we implemented the algorithm for tetrahedral meshes and also designed and implemented a visualization tool for looking at these structures.

We implemented the Morse-Smale algorithm to construct the descending 1-manifolds. We also designed and implemented a tool to help visualize these descending manifolds. The visualization tool was developed using C++ and VTK. It provides the user with the ability to visualize the input dataset as a mesh and as isosurfaces for different isovalues. The user can look at the entire set of descending 1-manifolds and filter them based on the function values. The user can optionally use a color map to code the function value on top of each of the above structures. We hope that this tool will help users to identify features in their datasets. It could be used as an exploratory tool.

A lot of additions can be made to the visualization tool. Looking at other substructures of the Morse–Smale complex might be useful for exploring the input dataset. We are working on computing and visualizing the ascending manifolds.

A Computational Design Tool for Microdevices and Components Used in Pathogen Detection Systems

Andrew Nonaka
University of California, Davis
Mentor
David Trebotich
CASC

Summary

Because of recent events including 9/11, the war on terrorism, and the war in Iraq, Lawrence Livermore has taken an increasing interest in biological pathogen detection. My summer research involves creating computational models for MEMS (microelectromechanical systems) devices for pathogen detection systems. A systemwide computational model will save time and money for LLNL scientists because they will have the ability to accurately predict if their devices will operate correctly prior to fabrication.

My objective for the summer is to become familiar with the software infrastructure currently in development to solve similar problems and to adapt the software to deal with biological flow through microchannels driven by a magnetohydrodynamic (MHD) pump.

The project will rely on a software infrastructure known as Chombo, which is designed to solve partial differential equations using finite difference methods on grids. The infrastructure EBChombo (embedded boundary) is an advanced version of Chombo equipped to handle the complex geometries found in MEMS devices. Both make use of adaptive mesh refinement (AMR) to determine which spatial locations in the problem require the most computations to solve accurately.

I am working with David Trebotich from LLNL and Greg Miller and Thomas Marshall from the University of California at Davis to reprogram the infrastructure to solve our problem. We are using a series of discrete hyperbolic and elliptic solvers to implement viscoelastic flow with MHD in Chombo. We have also learned how to program complex geometries using the EBChombo framework and have run two-dimensional examples of ideal polytropic gas dynamics using our geometry.

In the next few months, we plan on implementing a particle scheme to simulate proteins and other large molecules in the fluid. Then we wish to extend our algorithm for viscoelastic flow with MHD to EBChombo to handle more complex geometries. We also wish to implement the AMR algorithm to speed up our code. We will be able to verify our results with actual experiments being performed at UC Berkeley.

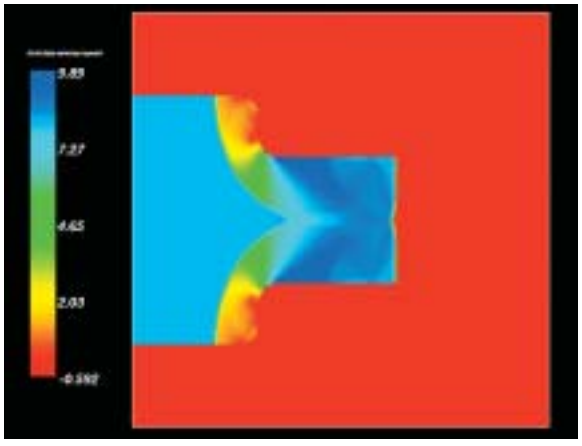


Fig. 1. Chombo output for an ideal polytropic gas. A shockwave of high-density and high-velocity polytropic gas has entered a contracting channel filled with low-density, stationary polytropic gas. Color scheme details the relative values of the x-component of fluid velocity.

NIF Automatic Alignment Archiving

Brian Overstreet
University of Colorado
Mentor
Punita Sinha
NIFE

Summary

The automatic alignment (AA) subsystem of Livermore's National Ignition Facility (NIF) laser system produces an enormous amount of data that must be analyzed by researchers to maintain system-critical components and learn about shots. The data must be archived to a database for easy analysis and not relegated to cryptic text files. The archiving mechanism developed to meet this need employs the languages Ada, CORBA, and Java to send the data to a database.

The AA control system was written in Ada, but Ada cannot communicate with the database because the database no longer supports it. Therefore, the decision was made to use Java's JDBC. The problem of interlanguage communication between Ada and Java came up, and the solution was to use CORBA's IDL. Another key issue was understanding the data. The AA system is large and complex and has many components. Once the data were understood, the next step was to draw an entity relationship diagram to design the database tables. With understanding of the data in place, the structure of the IDL was made to reflect the data. Then the Ada code was written to pull the data out of AA by putting it into the structure defined by the IDL. Finally, Java took hold of the CORBA. It unpacked any data type that the IDL sent it and repackaged it in an SQL statement to insert into the database. At the completion of this project, operators can archive the data with a single mouse click. Once the archiving process is complete, researchers can pull up a web tool to query the database and download the images associated with AA.

Future plans include writing a tool to link the multiple tables associated with automatic alignment into an easy-to-use interface.

Computational Gene Annotation and Human–Chicken Comparative Analysis

Samir Pandurangi
University of California, Davis
Mentors
Ivan Ovcharenko
CAR
&
Art Kobayashi
EEBI

Summary

This project involves a comparative study of the human and chicken genomes. The focus of the study will be on human chromosome 19 (Chr 19). This chromosome is unique in that it is relatively small yet contains a disproportionately large number of genes. Chromosome 19 in humans is closely associated with chromosome 28 (Chr 28) in chickens. The chicken organism turns out to be an ideal candidate for study because of its relative position in the evolutionary tree with respect to humans. Because of evolutionary conservation, the human and chicken genomes share many similarities.

Most of the chicken bacterial artificial chromosomes (BACs) in this study originate from chicken Chr 28. The first step of this study will involve annotating chicken genes—that is, using BLAST results to locate where genes possibly lie within the BACs. It is important to keep in mind that we are developing a technique for the prediction of coding structures in novel genomes. This method will make it possible to annotate any genome without a need for full RNA transcripts or heavy experimental evidence. These sequenced BACs homologous to Chr 19 comprise only a small portion of the chicken genome. However, if we are successful in annotating the genes that lie on these BACs, we can use our method to annotate any sequenced genome with very little experimental evidence. After annotation, we perform a comparative analysis of the chicken and human genomes.

The chicken genome has not yet been annotated. However, tools for gene analysis and annotation are plentiful. Some of these tools have been developed by organizations where researchers are conducting other annotation projects. One such tool, which is of central importance to this project, is the University of California at Santa Cruz (UCSC) genome browser. This is an online tool that helps to visualize genome characteristics. The browser that we installed locally was originally developed at UCSC (UCSC Human Genome Browser) to accommodate the needs of visualizing and annotating the human genome.

We have decided to use the framework of this tool to construct our own online browser to visualize the chicken genome. This involves a modification of the current tool provided by UCSC, an open-source C code. This modification is not trivial. The package involves tens of thousands of lines of source code with multiple, interdependent binaries and little documentation. I first had to understand the code, modify it to suit our needs, insert some of my own C code, and write other programs to interface with the existing binaries (see diagram).

After building the browser, I annotated the chicken BACs. Our browser currently has a wealth of information about protein sequence similarity to our chicken BACs through the many BLAST alignment hits. I use protein-similarity-driven gene predictors (e.g.,

Continued

Samir Pandurangi
University of California, Davis

Summary continued

GeneWise) to take these sections of similarity as input and output a set of predicted genes. In order to capture regions of the DNA that do not have any BLAST-produced protein similarity, I run other gene predictors (e.g., GenScan) that are indifferent about whether there is existing protein similarity in that region or not. This sheds some light on possible gene sequences between regions of protein similarity.

After completing annotation, my group and I will continue the project with a comparative analysis of the two genomes, using many of the same ideas and techniques developed in earlier parts of this project.

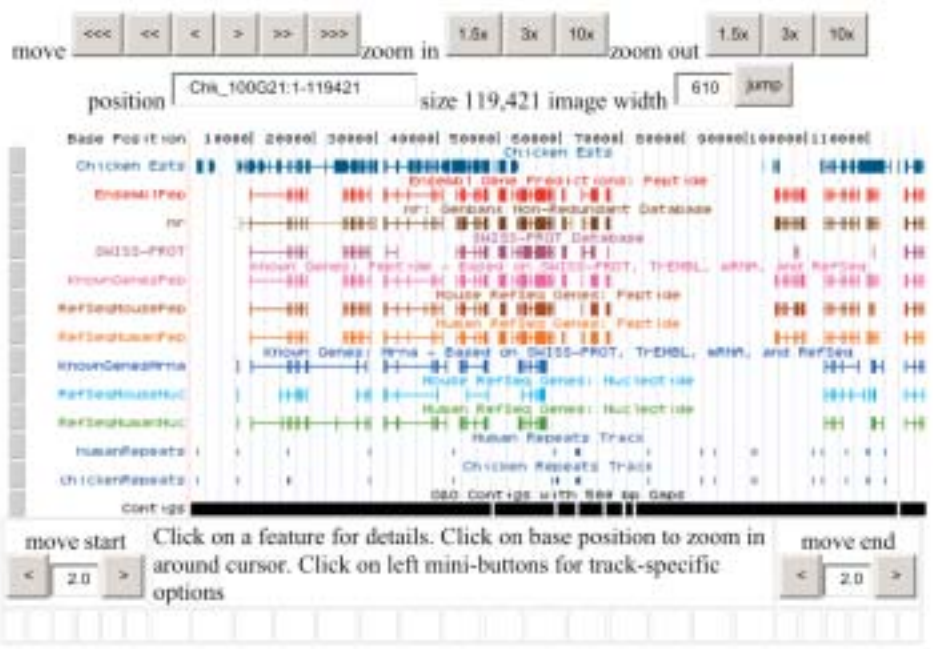


Fig. 1. UCSC Genome Browser on Chicken BACs

Memory Access Characterizations of Scientific Benchmarks

Ricardo Portillo
University of Texas, El Paso
Mentor
Jeffrey Vetter
CASC

Summary

The constant increase in raw processing power of contemporary chips has led to an improved performance on the computational side of a program's execution. And yet, the fact that performance in memory-access times and memory bandwidth has failed to keep up with this rate of improvement means that an application's overall performance is often hampered by bottlenecks in memory. To reduce these bottlenecks, efforts have been made to develop memory hierarchy systems that take advantage of memory-access patterns of applications. Further improvements can be achieved if memory systems are designed with access patterns of specific types of applications, such as scientific programs, in mind. To do this, metrics on how an application will access memory on a given hierarchy must be gathered in order to choose the most appropriate system for that application.

To gather these metrics, we have modified a version of the sim-outorder processor simulator, which is part of the Simplescalar suite of simulation tools. Sim-outorder is capable of simulating the out-of-order pipelined execution of an application running on a given architectural configuration, including the memory hierarchy, and produce a large range of performance statistics. Currently, sim-outorder supports the Alpha, PISA, and 32-bit PowerPC instruction set architectures. The most important modification to sim-outorder that we have made enables it to produce a detailed address trace of all the memory locations that an application accessed during execution and in the order that they were accessed. Using this data, we are now in a position to gather important metrics such as memory reuse distance and working set size of a program's execution that might enable computer architects to design complementary memory-hierarchy structures. In addition, these metrics may allow software developers to better optimize their code for existing memory hierarchies and therefore achieve better performance.

Because processor simulators such as sim-outorder may take hundreds or even thousands of times longer to execute than running the tested program natively, we have also added features that will allow us to simulate only small portions of a programs execution and thereby reduce the waiting time for simulation statistics. In addition, we have enabled sim-outorder to produce metrics on the instruction mix of a given program running on the simulated architecture. This information will give us more insight into what types of instructions are most used in a given execution and modify our simulated architecture configuration accordingly.

Computational techniques to extract meaningful metrics from the address trace produced by sim-outorder are now in development. Once these techniques are implemented, a study will be conducted on the memory access patterns of specific applications that are relevant to national security, including those which form part of the ASCI Purple suite of scientific benchmarks.

Out-of-Core Surface Parameterization and Remeshing

Serban D. Porumbescu
University of California, Davis
Mentor
Mark Duchaineau
CASC

Summary

With increased computing power, improved computational simulations, and improved sensor technology come vast amounts of information. To use this information, scientists must be able to visualize and manipulate this data in real time. Often, this data is too large to fit in the main memory of high-end workstations and, in some cases, even in the main memory of the supercomputers that generate the data.

Our continued research has focused on the development of a robust and efficient out-of-core surface remeshing and parameterization technique. These parameterized surfaces are amenable, among other things, to continuous-level-of-detail control, multiresolution editing, texture mapping, and, most importantly, compression.

Our technique begins with an arbitrarily triangulated surface. The end goal is to efficiently transform this surface into one that represents the same exact shape but is instead composed of very regular and similar-looking quadrilaterals (like those on a piece of graph paper).

Our past work has lead to an algorithm that constructs a hierarchy of surfaces that map to one another and that can be easily traversed. Our current work focuses on minimizing the distortion and maximizing the similarity of the quadrilaterals that comprise the surface.

We use a technique that is loosely based on the physical simulation of damped springs and masses coupled with the notion of a best-fit square (in the linear least-squares sense). The best-fit square represents what the perfect universe would look like to its respective quadrilateral if it were distortion free. These best-fit squares in combination with the spring simulation force distorted quadrilaterals to transform into similarly sized and shaped squares.

The algorithm is unique in that all operations require only local information. This translates into a technique that is easily made to work both out-of-core and in parallel.

This technique has been designed to work out-of-core, but the exact out-of-core algorithm needs further study and development. Our current distortion metric shows promise but needs further enhancement to ensure the resulting surfaces are manifold.

BlueGene/L Research Project

Daniel A. Reddeg, Jr.
United States Naval Academy
Mentor
Kim Yates
CAR

Summary

My research project at the United States Naval Academy focuses on supercomputing—its history, current uses, and where it is going in the future. My objective during this internship was to gain new insight into the world of supercomputing by learning as much as possible about the BlueGene/L project, as well as the many other supercomputing ventures taken on by LLNL. To this end, I sought to gain as much hands-on interaction with the different software and testing mechanisms employed in developing the new BlueGene/L system as possible.

My study of the BlueGene/L system was aided mainly by the BlueGene/L team, and their assistance in providing me with all of the materials for my research was greatly appreciated. In addition, I was able to test specific developments of the BlueGene/L system on the simulator, which provided me with raw performance data that, more than anything, verified the information I had received from the BlueGene/L team.

From the research I have done here, I plan to return to the Naval Academy and prepare a technical paper and presentation to submit to the research board at the Academy for review and approval of my final graduation requirements.

Data Visualization Project

Eric Scamman
University of California, Los Angeles
Mentor
Scott Brandon
DCOM

Summary

The purpose of this project was to assist Livermore physicist Scott Brandon in his validation and verification work with the Kull simulation code by providing visualization tools. The first tool needed was one to present multidimensional data in a visual format. By doing so, a large number of parameters could be compared at once to aid in error-limit estimation. The second tool's purpose was to display one-dimensional slices of data through the mesh used in a Kull simulation during and after a run. To be successful, it needed to be easily incorporated in the input deck for a Kull simulation.

The concept of the glyph was used to represent multidimensional data. Glyphs are small objects that can have many attributes and can be plotted over an *x–y* plane. Vectors in a vector field are examples of simple two-dimensional glyphs. Eventually, five-dimensional glyphs were employed, consisting of two vectors each and a color. Seven dimensions were successfully displayed with the plotter, which included options to filter parameters for certain value ranges along with other editing tools. The application was scripted in Python with the use of the PyGist, Tkinter, and Numeric modules.

The second tool was also constructed in Python using PyGist and Numeric. To perform a successful slice, the first objective was to build a line. This line consisted of a list of points separated by an increment of distance determined by the user. The essential function of the tool was to determine which zone each of these points lay in. The approach of breaking the mesh space into cubes allowed for fast searches over nearby cubes to find the point-containing zone. PyGist allowed a simple plotting medium to display the constructed line. Finally, code was added to the program to allow it to run over multiple processors in order to be used for very large simulation meshes. Upon completion, the code was successfully used to display plots of data vs. radius in the Sedov Blast Wave Test Problem over various user-defined lines.

The future of this project consists of the implementation of the above tools with Kull simulations along with improvements to the tools as necessary. Such improvements could include rewriting some Python routines in C and editing for more efficient scalability.

XS4C: Automatic Code Generation for Complex Step Method

Andrei Schaffer
University of Iowa
Mentor
Radu Serban
CASC

Summary

XS4C is an automatic code generation tool that parses a C user code and generates the C++ complex arithmetic version of that code. XS4C computes derivative information (gradients, Jacobians) via the Complex Step Method and is used in conjunction with LLNL's Sundials ODE/DAE numerical integration package. Although XS4C was initially designed for CVODE's integration package, which is part of the Sundials package, it is now a more generic tool for generating complex arithmetic C++ code. The complex arithmetic code that is obtained uses an enhanced version of the complex class provided by C++ Standard Template Library (STL). The complex code uses complex objects for scalars, and it duplicates arrays by providing a real-part array and an imaginary-part array, which are manipulated as blocks whenever possible for efficiency reasons.

XS4C parses the C source code using the C++ libraries provided by ROSE, a parsing tool developed at LLNL, and the complex arithmetic code is generated by attaching real and imaginary arithmetic code (expressions, declarations, function calls) attributes to each node in the Abstract Syntax Tree (AST), which was obtained as a result of the parsing operation. As a result, a set of C++ output files is obtained. These files are compiled and linked with Sundials libraries and with the user code, through a wrapper that is partially static and partially dynamically generated by XS4C. The result is an ODE/DAE numerical integration solver that obtains its necessary derivative information via Complex Step method by using the C++ routines generated by XS4C. Experiments performed so far show an increase in the accuracy of the solution compared to the finite difference computation of derivative information. This increase is because Complex Step method has the advantage over finite difference methods of avoiding catastrophic cancellations that usually accompany subtraction operations when step size is close to the round-off unit. As a result, Complex Step method can use step sizes as small as the round-off unit as opposed to step sizes that are typically proportional to the square root of the machine's round-off unit for finite difference methods—hence the improvement in the accuracy. It should be noted that the Complex Step method error is of order two when the step size is small enough. Therefore, if a step size proportional to the machine's round-off unit is used, the derivative information is as accurate as the analytic derivative computation performed on that machine.

The documentation and manual of XS4C are tasks that remain to be done. The second version of XS4C will provide a more efficient complex code by complexifying only a smaller set of variables and functions, namely only those that are in the intersection of program slice containing variables or functions that depend on the input variables (the independent variables, i.e., variables with respect to which derivatives are computed) with the program slice consisting of variables or functions that determine the output variables (the dependent variables, i.e., variables that are differentiated).

PFD*Now*: Desklog Module

Jonathan Schifman
University of Southern California
Mentors
Yousseff Abed
&
Priya Basu
EEBI

Summary

The duties of LLNL's Security Information Technology & Engineering (SITE) Group include developing information systems that enhance the security of the Laboratory and the efficiency of the Safeguards & Security Department. PFD*Now* is a web-based administrative system designed to automate, organize, and streamline the workflow processes that occur within the Protective Forces Division (PFD). The system involves several modules dealing with personnel, inventory management, officer scheduling, daily reporting, and workflow. The PFD*Now* system has been in development for over one year and is being used on a per-module basis.

Beginning in May 2003, the Desklog module became the highest priority item within the project. The Desklog module is designed to handle all information, workflow, and tracking regarding PFD incident reports, Central Alarm System (CAS) event logs, and shift logs for operations sergeants. By automating these tasks, all information is in a centralized location, which is accessible to any user with appropriate permissions. Desklog's main objective is to serve as a catalyst to daily PFD operational issues and track workflow in an easy to manage system.

The PFD*Now* application is a web-based Java application using a Model/View/Control architecture. All code was written using a framework authored by PFD*Now* architect Sean McFadden. The MVC framework uses JavaBeans, XML, and JDBC, interfacing with an Oracle 9i database, along with a single controller servlet, which handles server-side logic (model and control). The user interface (view) is presented using JSP pages. Priya Basu, Sean McFadden, and I met with PFD administrative personnel Clea Marples and Marlene Dutchover for most of the requirements analysis. All incident reports were previously paper forms, and event logs were completed using Microsoft Excel.

Each incident report was handled separately, but similarities between certain ones allowed us to design with an abundance of code-reuse techniques. The appropriate form fields were evaluated, as was the user interface and use case specifications for each form. Some reports facilitated special features, such as personnel assignment searches and calendar widgets for date and time selection. By mid-August, we were able to complete the automation of every major incident report in use by PFD.

The other main aspect of the Desklog involved the automation and integration of two disparate systems: the CAS Desklog and the Operations Sergeant Desklog. Much of the information recorded in these logs was redundant and caused unnecessary work to be done by both the operations sergeants and CAS operators within PFD. We were able to use a stripped down version of our base incident report form implementation to handle events. In the final system, the incident report Desklog and two event Desklogs have been unified in a single, easily manageable interface.

Full integration of email-based report distribution, notification, and reporting is currently in development. In addition, customized user alerts are the only other crucial component needed to begin user testing and training with the Desklog module of PFD*Now*.

MObIUS (Massive Object Integrated Universal Store)

Jennifer K. Sirp
California State University, Sacramento

Mentor

Terry Brugger
NAIC

Summary

General frameworks for distributed computing are slowly evolving out of Grid, Peer architecture, and Web Services. The following are results of a summer-long survey into distributing computing practices: First, Legion and Cactus-G have achieved the most in terms of providing an all-purpose application environment; second, extending a local programming environment to operate in a highly distributed fashion can be facilitated with toolkits such as Globus; and third, building a new system from the ground up could be realized, in part, by using some of the following components—an object-oriented database, Tapestry, JXTA, BOINC, Globus, component architecture technology, XML and related libraries, Condor-G, Proteus, and ParMETIS.

The objective of this research was to locate systems that could potentially contribute to the development of MObIUS, a general, distributed, object-oriented, framework. In contrast to current implementations of Web Services, Peer-to-Peer, and computational Grid middleware, the proposed system should be able to efficiently handle a wide variety of applications and adapt to a dynamic environment.

The approach taken was to first assess the state of distributed computing today. Next, systems both large and small were explored. Because time was short and it was unclear whether or not the proposed system existed, a broad approach was taken first. Later, the most relevant systems were reevaluated and questioned.

Distributed computing is an exciting topic; object-oriented technology and supercomputing are being reflected in this field of study. The whole concept of using the collective power of the Internet is fascinating. I am interested in seeing how advancements in component-based architecture, agent-oriented programming, and artificial intelligence will affect distributed computing in the future.

Data Visualization and Programmable Graphics Hardware

Jacob Stevens
Northern Arizona University
Mentor
Dave Bremer
ICCD

Summary

Recent innovations in hardware graphics technologies allow the traditionally fixed functionality of the graphics pipeline to be customized with short assembly language programs that are executed directly on the graphics processing unit (GPU). The objective of this study was to determine whether this new feature has the potential to benefit data-visualization techniques.

Several proof-of-concept programs were written that implemented key data-visualization techniques, such as paletted volume rendering, isocontouring, and simultaneous rendering of multiple volumes (see Figure 1). It was found that using programmable graphics hardware has the following notable benefits for data visualization, especially in the area of volume rendering:

- intensive per-fragment (pixel) computations are executed on the GPU, freeing the central processing unit (CPU) to perform other tasks,
- as much as 75% less video memory is required to render single-paletted volumes,
- fifty percent less video memory is required to render dual volumes,
- specialized GPU programs can be written to provide backward compatibility for older hardware operations that may be eliminated from future devices,
- when changing visualizations, only the currently active program on the GPU must be changed, rather than reprocessing an entire data set, thus allowing different visualization techniques to be explored interactively, and
- GPU programs have direct access to the real-time graphics application programming interface state, such as OpenGL (http://www.opengl.org), allowing visualizations that use techniques such as view-dependant lighting.

Overall, the use of programmable graphics hardware was encouraging, and although the technology has a tendency to be somewhat unstable and unpredictable because it was introduced relatively recently, it is likely to become an important tool in future visualization projects.

Because of the flexibility of using GPU programs to display data, an interface is being considered that will allow end users to have high-level language access to the functionality of programmable graphics devices. This will eliminate the need for assembly language programming and also provide the user with a rich set of useful mathematical operations commonly used in visualization.

Continued

Jacob Stevens
Northern Arizona University

Summary continued

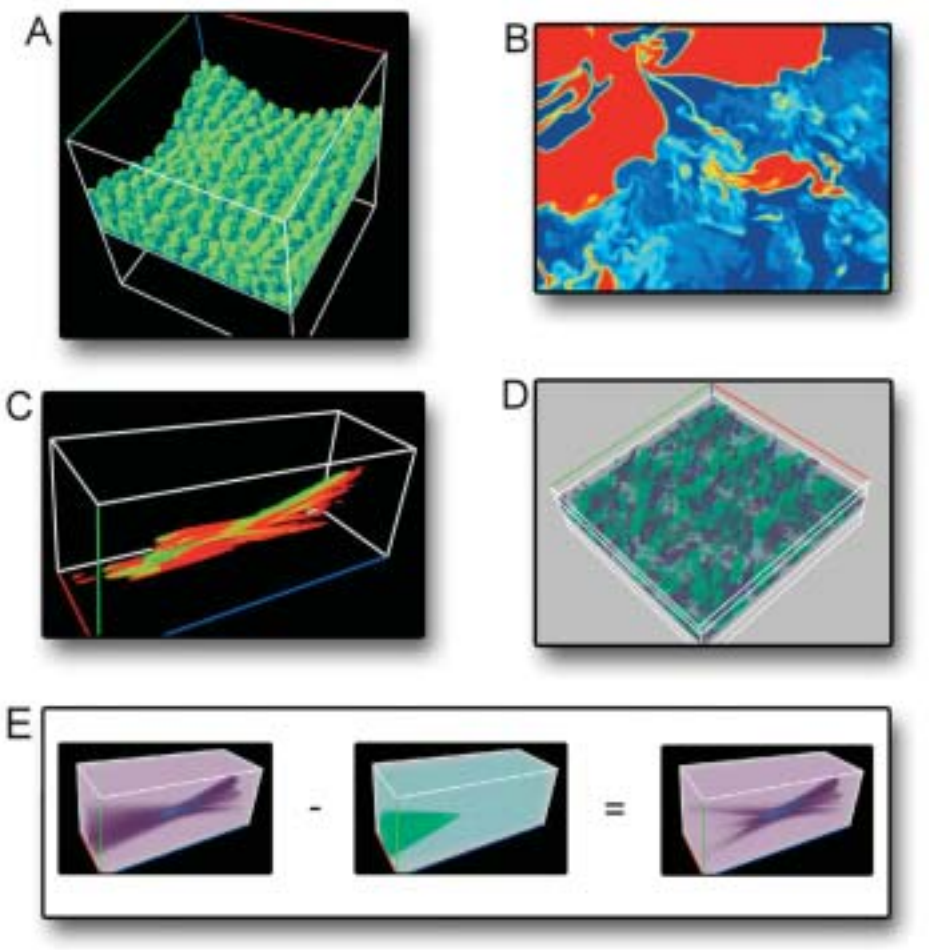


Fig. 1. Several volume renderings computed using programmable graphics hardware: (A) shaded volume rendering, (B) volume rendering with user-defined color palette, (C) gradient of data visualized as hue, (D) simultaneous display of multiple volumes, (E) point-wise subtraction of one volume from another.

Porting an Isosurface Calculation and Visualization Package from Yorick to Python, Utilizing SWIG

Mark Stuppy
University of Missouri, Rolla
Mentor
Steve Langer
DCOM

Summary

To increase the possible user base of an isosurface package written by Steve Langer, the decision was made to make it available in both Yorick and Python. These two scripting languages are functionally similar enough to allow the package to be used in basically the same way in both versions, but also different enough to give each version advantages and disadvantages. The goal of this summer project was to see if this addition could be done easily while adhering to set design goals and if so, to determine a method for doing so. These design goals are as follows:

- retain the same user-end functionality in the Python version as in the original,
- retain the same memory-management features in the Python version as in the original,
- complete the conversion with the least amount of scripting language specific changes to the original, compiled code as possible, and
- if possible, use the object-oriented nature of Python to extend the package's functionality through the use of object-member functions.

The project was first divided into two main goals: first, port the calculation functions into Python and second, port the visualization functions into Python.

For the first part, a code-integration package called SWIG was used to help wrap compiled C functions to be called from Python. SWIG's main job in this project was to automate the process of converting Python variables and classes into C-style variables and structures to be passed into compiled C functions. This process can be done by hand on a per-function basis, but in the past that has been shown to take a long time and to be quite prone to error. SWIG, however, only requires that developers write the upper-level code for converting object types or specific arguments, and then it generates the lower-level memory-management code and uses it wherever necessary. After wrapping some functions of this type, all that was left for the first part of the project was to mirror the structure and code of all of the associated Yorick functions in Python.

The second part of the project dealt with getting the same visualization functionality available in Yorick into the Python version. This required much more than simply wrapping more functions. It was also necessary to restructure some of the original code to be less dependent on certain Yorick-specific functions and therefore more interoperable with both languages.

The next step in this project is to add the multiprocessor calculation functionality created last summer by the author into the new Python version of code. This should be quick and painless; however, time constraints did not allow it to be completed this summer.

A Nodal Approach to Algebraic Multigrid (AMG) for Systems

Ryan S. Szypowski
University of California, San Diego
Mentor
Ulrike Yang
CASC

Summary

Algebraic multigrid (AMG) is a linear solver applicable to a wide range of problems. However, when the linear system to be solved is derived from a system of coupled partial differential equations (PDEs), AMG is often unsatisfactory for use as either a stand-alone solver or a preconditioner.

In such cases, it is necessary to augment the standard AMG scheme using knowledge of which unknowns are associated with a single node in the discretized PDE. One approach, known as the unknown approach, is to ignore connections between the different functions and coarsen them separately. Another approach, known as the nodal approach, is to categorize nodes of the discretization, as opposed to unknowns, as coarse or fine nodes, thereby keeping the coupled structure through all levels of the multigrid hierarchy. The nodal approach was studied and its performance as a preconditioner was compared to standard AMG and the unknown system version.

A large amount of code was added to a serial version of the Hypre library to accommodate block-structured sparse matrices. Also, special interpolation and relaxation routines were developed for use with the block sparse matrices.

When tested as a preconditioner for conjugate gradients on various two- and three-dimensional linear elasticity problems, the nodal AMG converged in fewer iterations than standard AMG. When compared to the unknown approach on simple problems, the nodal version was comparable, sometimes converging faster and other times slightly slower. However, on more difficult elasticity problems, the nodal version converged in far fewer iterations than the unknown version.

Although the results are promising, more theory and experimentation is required to derive better interpolation methods for nodal AMG.

Identity Access and Management Service

Robert Blake Taylor
Northern Arizona University
Mentor
Richard Mark
ICCD

Summary

The purpose of the project is to develop a new Identity Access and Management Service administration tool, replacing the current DCE tool, because of the deployment of a new security infrastructure for ASCI platforms and services within Livermore's Computation Directorate. The new administration tool must support a model that decouples the GUI from the back-end server, is platform independent, and has supporting libraries or components that allow for easy development within a distributed system. The prototyping of a new tool will also allow for the implementation of the necessary and requested changes within the current GUI design.

In developing the administration tool, our group met with the current DCE tool users to generate a list of the design changes to be implemented in the new GUI. The implementation of these design changes in the GUI has greatly simplified the process of completing what were previously complicated tasks. The GUI also implements the use of mnemonics and keyboard shortcuts, greatly increasing the speed at which users of the tool can complete the necessary tasks in updating user and group information.

Our group also focused on the task of designing the new system model that decouples the GUI from the back-end server. The new system model requires a way of sending data between the GUI and the back-end without tying the implementation to a single language. XML, which is language independent, was implemented as a means of storing the current state of this data for exchange between the GUI and the back-end.

The GUI currently relies upon the easy creation of serialized XML encoded user-defined structured data types, such as a User or Group, to update and change user and group information. Serialized types are not compatible between versions of the J2SE and therefore cannot be used in the final implementation of the GUI. The implementation of SOAP, a XML-based object-access protocol, is currently being researched as an alternative to using the serialized data types.

Implementing the SOAP protocol will allow the GUI, a SOAP client, to invoke procedure calls for the access and modification of user and group information on the back-end server via SOAP services located within the intermediate server. The implementation of a SOAP client and server will provide the flexibility required to allow the GUI or the back-end to be replaced or modified at any time.

Scientific Visualization Movies

Peter Tipton
University of Southern California
Mentor
Hank Childs
DCOM

Summary

After a physics simulation has been run and its output has been stored, the scientist who ran the simulation wants to see what happened. The process of actually looking at computer data from a physics simulation is called visualization.

Sometimes scientists want to see how the simulation progressed through time. They want to see a movie, which sequentially shows each time step of the simulation and displays whatever information is desired. This is a visualization movie.

Using the visualization tool VisIt, we can read in the data from a physics simulation and make a picture from it. We can show different aspects of the simulation, and we can show the simulation from different views. However, we can't make a movie in this manner.

To make a movie with VisIt, we have to run it from its command line interface. VisIt comes with the ability to make movies, but only with a scripting language. VisIt uses Python to do its scripting. By typing in various commands, we can tell VisIt what we want to see, where we want to see it, and when we want to see it. We now have control over the temporal aspect of a movie. We can change the view of the simulation throughout the movie, giving us a different perspective of things. We can show the outside of something and then reveal what is inside. In this manner, we can animate anything we want.

The other aspect of making a visualization movie is being able to show several smaller movies at once. We want to show one aspect of a simulation made with VisIt while simultaneously showing a different aspect of the same movie. We want to make a bigger movie that is a split screen of two smaller movies. This can be done with Unix Shell Scripting.

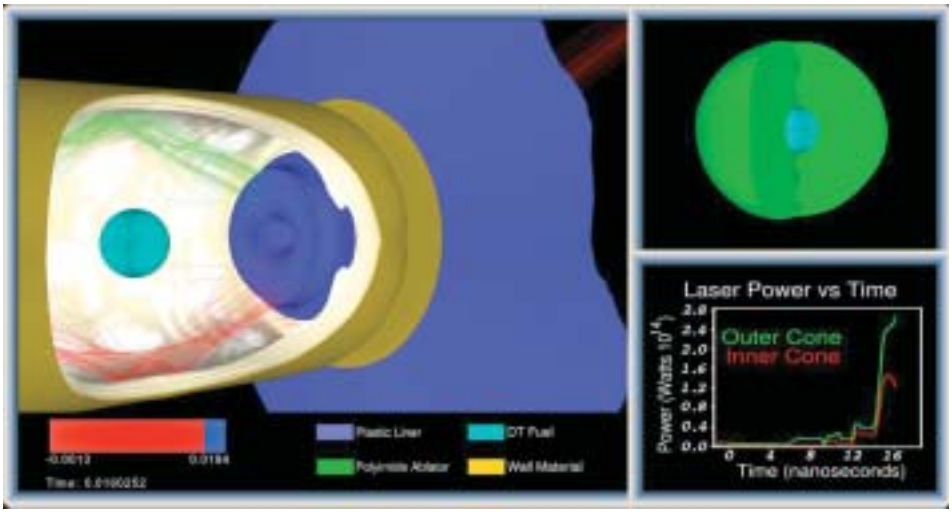
If we have two image sequences (two movies), we can issue various resizing and compositing commands from a Unix shell script that will go through all of the frames of the movie and composite them into one large movie.

This past summer, I used Python scripting along with Unix Shell Scripting to make movies in this manner. I have finished one movie for Michael Marinak, who is visualizing an indirect drive laser-fusion simulation of the National Ignition Facility (NIF), showing the ignition of the target. (On the following page is a picture from this NIF movie.) Having made this movie, I have learned a great deal about scripting, Unix, and parallel computing.

Continued

Peter Tipton
University of Southern California

Summary continued



Visualization of a simulation of laser-fusion ignition of an indirect-drive target on the National Ignition Facility.

Run Time Adaptive Load Balancing for Large-Scale Parallel Computing Systems

John Viles
Stanford University
Mentor
Jeffrey Vetter
CASC

Summary

The objective of my summer assignment was to evaluate the conceptual viability of a run-time load-balance control system for large-scale parallel applications. Effective load balancing is a difficult problem to solve statically and is especially difficult to manage in the face of variations in the run-time environment during the execution of a large job.

The goals for an effective run-time load-balancing scheme include:

- distributability and scalability with low-run-time complexity,
- adaptability to changes in the computing environment,
- monotonicity toward improved parallel program performance, and
- automatic implementation, requiring minimal programmer effort to effect.

Accordingly, the load-balancing control system developed in this work is designed to augment a general parallel program by providing adaptive self-balancing capabilities with a minimum of programmer or user intervention. The augmented executable code becomes capable of adaptively optimizing its own use of parallel resources in order to run as efficiently as possible and to facilitate maximum parallel scalability.

A working prototype of the control system was successfully implemented in Matlab and tested by simulation with several different algorithms for systems of up to 64 processors. In these trials, the control system was shown to be robust and effective at updating the computational load distribution in response to both normal variability (up to 10% random degradation) in underlying system performance characterized by effective processor speeds and communication bandwidth and latency and sudden, substantial changes (up to 500% degradation) in these parameters.

To be feasible, the load-balancing control system needs to be mathematically separable into one set of components that depend on the parallel algorithm to be optimized and another set of components dependent only on program input parameters and hardware performance attributes. The former can be generated statically at compile time, while the latter must wait for run-time instantiation.

The prototype developed this summer at LLNL automatically generates a separable performance model, consisting of one set of statically generated, machine-independent formulas that capture the time and communication complexity of the program, and a second set of dynamically evaluated architectural parameters and transformational operators. As measurements of processor and communication timings update the system parameters during program execution, the control system employs a parallel version of Newton's algorithm to reduce variations in parallel execution times across the processor domain. The result is a flexible system with low run-time demands on hardware resources.

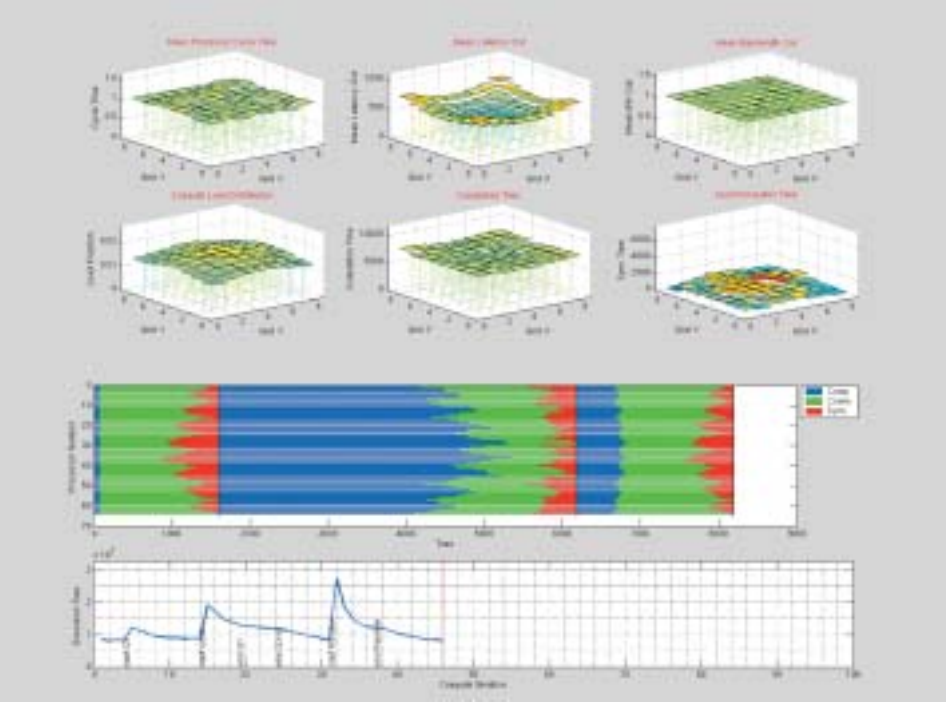
Continued

John Viles
Stanford University

Summary continued

Future work needed to support this project includes:

- instrumentation of MPI function calls at a low level to gather timing information without the intervention of the main user program,
- further analysis of the separable computational model to identify where it is applicable,
- implementation of a feedback control scheme to allow the control system to operate with a reduced fidelity performance model,
- extension of the performance model to manage functional parallelism effectively, and
- development of static program analysis tools to characterize run-time complexity based solely on either the source or intermediate representation of the program.



The effect of the load-balance control system on the execution time of a sequence of iterations of the conjugate gradient algorithm on a 8 × 8 (64 processors) grid. The hardware simulated includes a variation of 0–10% (“noise”) in communication latency and bandwidth and in processor throughput. The trace at the bottom of the figure illustrates the effect of three different major system disturbances on execution time and the recovery of the program as the load control system shifts computation to compensate.

Network Vulnerability Assessment Project

Qian Wang
Massachusetts Institute of Technology

Mentor

Terry Brugger
NAIC

Summary

The Network Vulnerability Assessment Project is the task of many computer security professionals under the Information Operations Assurance Center (IOAC) at LLNL. The project involves using tools, known as Deckard and MMM, to collect as much data as possible about a particular network. The tools create XML files of this data, which are read by the loader. The loader loads the data into the Network Mapping Database (NMDB) so the Graph Viewer can query and display the data in a user-friendly form. Computer security personnel can then use the Graph Viewer to look for vulnerabilities in the network.

Different computer security professionals work on the different aspects of the project. I helped improve the Deckard tool by creating a way to automate queries into the domain name database and to store the retrieved information. I also created two tools to monitor network traffic of an IP address and to detect the operating systems of machines based on packet information sent by such machines.

I learned the Python programming language to develop the DNS database text parser. Given a domain name, the parser automatically performs a Dig query. For each type of query, it reads in the lines produced by Dig and stores them into a file. Then it parses through the lines to determine whether the name provided is a valid domain name, if it is an actual domain name or just a single machine, the IP addresses of the domain, and the name and mail servers associated with the domain name. It stores each resource record into a list and makes sure there are no duplicates.

The fingerprinting tool uses the Disco fingerprint database to determine the operating system of a packet sender. However, it stores the fingerprint database information into a tree and then searches through the tree recursively to find matches. Disco searches through the database linearly and thus is not as time efficient as the fingerprinting tool. In each run of the tool, it creates the tree, listens for SYN/ACK packets or reads packet information from a file, compares the packet information to information in the tree, and then stores the possible operating systems of the sender into a vector. The fingerprinting tool ran about twice as fast as Disco. It can also find multiple possible matches for the OS, whereas Disco stops the first match it finds.

The tool can be vastly improved by incorporating some active fingerprinting techniques. NMAP searches through its active fingerprinting database linearly. That information could be stored in a tree to improve the speed of searching through the fingerprint database. I could use passive fingerprinting techniques to narrow the possible operating systems down and then only perform the active tests on the remaining systems. This would reduce the number of active tests needed and thus make this tool stealthier and more efficient than technology currently available.

Authentication Transport System

James Waslo
Northern Arizona University
Mentor
Terry Brugger
NAIC

Summary

The primary motivation for the creation of the Authentication Transport System (ATS) was to improve upon an existing authentication system. This system was necessary to support use of one-time passwords through the RSA SecurID system on nonsupported operating systems. The commercial system provided by RSA does not support all operating systems used by LLNL and the Computation Directorate. It was therefore necessary to create an interface between those systems and the RSA software. The existing system, DRAT, was fully functional, but with room for improvement. The DRAT system consisted of two essential components, a server and a client library. It used a custom encryption scheme to provide secure communications. The server ran in parallel with the RSA software. The client library allowed services on a nonsupported OS to use the OTP system. The goal of AST was to maintain this general structure, but create room for expandability and more reliable security.

To accomplish this goal, many of the elements of the software had to be redesigned. First, and most difficult, was the implementation of secure communications. It was decided to use OpenSSL, a commercially reliable, open-source SSL implementation. Use of this package required additional education because none of the participating programmers had any experience with SSL or OpenSSL. Several small programs, such as an echo client/server pair were written in SSL to familiarize the programmer with the OpenSSL library. Second, the server had to be designed and written to implement the desired feature set. Rather than completely create a new server, the existing DRAT server was modified. Much of the original initialization and setup of the server was kept. However, the core operations of the server were replaced entirely. This included changing the encryption to use OpenSSL and handling authentication requests. Authentication requests were handled by type, beginning with a standard OTP authentication and including the possibility of future varieties of authentication types. Depending on the authentication type, a subprogram called an Authentication Module was called on to handle the actual authentication. The overall effect was smaller, lighter, and faster authentication.

There still remain some things to be accomplished on the ATS. The client-side library needs to be completed. This library will be used in recompiling common services, such as FTP or SSH, to allow them to use the ATS. Also, certificate and key verification must be added to both the server and client library.

Newton-Krylov Methods for Expensive Nonlinear Function Evaluations

Rebecca Wasyk
Worcester Polytechnic Institute
Mentor
Carol Woodward
CASC

Summary

Newton–Krylov methods have proven useful for solving large-scale nonlinear systems. An advantage of these iterative methods is that they only require knowledge of how the Jacobian acts on a vector but do not require storage of the system Jacobian. A difference quotient evaluated at each linear iteration is often used to approximate this action without slowing the convergence rate of the method. For systems with expensive nonlinear function evaluations, however, the requirement of a function evaluation for each linear iteration can result in a costly computation. The goal of this project is to explore convergence rates and time savings associated with using different approximations to the system function in the difference quotient.

Theoretical results developed by Peter N. Brown, Homer Walker, and Carol Woodward indicate that making a linear approximation of the most expensive nonlinearities in the difference quotient would result in a method with the same convergence rates as those of the unmodified method, where the full nonlinear function must be evaluated in the difference quotient. Since making a linear approximation could be expensive, we also tested a method where the most expensive nonlinearities were lagged, so that they did not need to be recomputed in the difference quotient at each linear iteration.

These two approximations schemes were tested on several problems using the KINSOL and CVODE solvers in the SUNDIALS program, a suite of solvers developed at LLNL for solving nonlinear algebraic systems, ordinary differential equation systems, and differential-algebraic systems. For most of the problems tested, the linear-approximation method did converge in the same number of iterations as were required when using the unmodified difference quotient. A small amount of time was saved when using this linear approximation on a few problems, but none of the test problems had expensive nonlinearities where we would expect to see the most time savings. Although using the lagged values of the nonlinearities was not supported by theory, the increase in the number of iterations required to reach a solution was not great for most of the problems tested. So, even though the number of iterations increased, in general, when using this approximation, time was saved while solving a number of problems, and these savings tended to be more significant than those that resulted from using the linear approximation.

Future work aims to develop some theoretical results on solving time-dependent problems using these approximations in the difference quotient. More tests are also planned on problems with more expensive nonlinearities, where the modifications made could really be beneficial in terms of run-time savings.

Deriving Prolongation Operators for Algebraic Multigrid (AMG) Using Compatible Relaxation

Dominique Wiest
University of Washington
Mentor
Charles Tong
CASC

Summary

Iterative methods are an accepted way of solving systems of linear equations with many unknowns. As these systems grow in size with the advent of parallel computing, scalable methods such as algebraic multigrid (AMG) have become increasingly important. Recent advances such as compatible relaxation promise to expand the scope of problems AMG can reliably be used to solve. One such type is solving nonsymmetric linear systems. We have looked at deriving new prolongation operators for such systems based on the improved coarsening provided by compatible relaxation. We compare these new prolongations on several convection–diffusion problems.

The AMG has three distinct parts: smoothing, restriction, and interpolation (prolongation). For nonsymmetric operators, the relationships among these are more difficult to analyze. For a symmetric positive definite operator, restriction and interpolation can use the variational property, which is easy to implement and commonly used. Focusing on the interpolation operator, we decided to examine the performance of an interpolation operator based on the idea of compatible relaxation (CR).

CR is a modified relaxation scheme that keeps the coarse level variables invariant. Compatible relaxation is used to determine coarse grid variables independent of any geometric information from the physical system. The convergence rate of CR can be used as a general measure for the quality of the set of coarse grid variables. In short, CR is based on the assumption that error after smoothing is small on fine-grid variables.

Since compatible relaxation can be viewed as a measure of coarse-grid quality, it is worthwhile to explore using CR to form the interpolation operator. However, CR theory is based on symmetric positive-definite systems, and it has not been proven to be a good measure on nonsymmetric systems.

However, the convection–diffusion problems investigated this summer using a compatible relaxation-based interpolation operator outperformed the standard interpolation operator found in Ruge–Stuben AMG. In some cases, it enables convergence where AMG had previously failed.

These preliminary results suggest that compatible relaxation is a viable means to improve the interpolation operator for these types of convection-diffusion problems. Future work needs to be done to examine other kinds of convection-diffusion problems and also using different smoothers in the compatible relaxation process.

ISCR

Institute for Scientific Computing Research

ISCR Sabbatical Program

ISCR Sabbatical Program

The ISCR hosts several sabbatical visitors each year. These faculty stay for 3 to 12 months and work on research in close collaboration with other LLNL staff members. Following are reports on sabbatical activities during FY 2003.

Distributed Relaxation Methods for Stokes Problems

Constantin Bacuta
Pennsylvania State University
LLNL Contact
Rob Falgout

Summary

In the context of solving indefinite systems arising from Stokes or Navier–Stokes equations using multigrid methods, a real challenge for proving convergence results is the construction of an appropriate smoother. The distributed relaxation method for the incompressible Stokes problem is based on a change of variables that leads to a lower triangular system with Laplace operators on the main diagonal, for which multigrid methods are more suitable. We propose a finite element formulation of Achi Brandt's distributed relaxation method. Using this approach, we intend to construct an effective and robust multigrid method for solving Stokes-type systems and other elliptic PDE systems, which can be reformulated as saddle-point problems.

In our finite element discretization of the Stokes problem, we introduce a discrete change of variables and study the properties of the transformed problem. We show that under reasonable regularity assumptions, the transformed problem admits almost block lower-triangular form. Here, “almost” represents the fact that the variables are coupled only through boundary values and the truncation error. We pay special attention to the boundary conditions of the new discrete variables. To impose the coupling Dirichlet boundary conditions, we use a stable quasi-interpolant that brings the derivatives of the pressure variable into the space of the velocity variables. We also prove the coercivity of the operator on the main diagonal of the transformed problem. We are in the process of implementing an extensive set of numerical experiments to demonstrate the effectiveness of the new algorithm we propose.

In the future, we will consider the almost incompressible elasticity problem. It can be reformulated as a saddle-point problem, and hence we can use a similar change of variables to end up with an almost lower triangular matrix suitable for smoothing. Another direction could be solving Maxwell equations, also reformulated as a saddle-point problem, and finding appropriate distributive relaxation change of variables and appropriate discrete spaces.

Computation of Incompressible Newtonian Flows

Zhiqiang Cai
Purdue University
LLNL Contact
Charles Tong

Summary

The primitive physical equations for incompressible Newtonian fluid flows are the conservation of momentum and the constitutive law. The constitutive law relates the stress tensor to the deformation rate tensor and pressure, and it states the incompressibility condition. By differentiating and eliminating the stress, one obtains the well-known second-order incompressible Navier–Stokes equations in the velocity–pressure formulation.

A tremendous amount of computational research has been done on these equations, but they may still be one of the most challenging problems in computational fluid mechanics. This summer, we studied two issues that are concerns of scientists at Lawrence Livermore National Laboratory: coupling between the momentum equation and incompressibility condition and unknown outflow boundary conditions in some applications.

The velocity–pressure formulation of incompressible Navier–Stokes equations is a coupled second-order partial differential system. The coupling dramatically complicates numerical procedure. First, spatial discretization requires a stable pair of finite element spaces to approximate the velocity and pressure. (Neither the Q1Q0 and Q1Q1 trial spaces are stable and have to be artificially stabilized.) Second, saddle-point problems at each time step due to implicit or semi-implicit time integration are very expensive to solve. To overcome such a difficulty, a popular approach dating back to the 1960s is to use splitting (projection) methods. The splitting method decouples a system through approximation in time, and hence the solution loses accuracy. More specifically, pressure boundary conditions used in splitting methods are artificial and inconsistent, causing inaccurate approximation and giving rise to a numerical boundary layer. We fixed this problem by a modification of the pressure calculation. An alternative is to use numerical methods, which we are developing, based on the stress–velocity formulation. Time integration of this formulation leads to a triangular system, which is decoupled. So we can calculate first the stress and then the velocity in a very efficient way.

The incompressible Navier–Stoke equation is an initial-boundary value problem in mathematical terms. This implies that one needs boundary conditions everywhere for its wellposedness. Unfortunately, many applications can be modeled by incompressible Navier–Stoke equations that have outflow boundary conditions that are unknown. This is a computational modeling issue rather than a purely numerical issue. Without physical intuition, one could use information from the differential equations themselves to approximate outflow boundary conditions. Nevertheless, it is impossible to get correct boundary conditions when using any traditional numerical approach. To compensate for incorrect outflow boundary conditions, one has to use a large computational domain. To tackle this problem, we are studying least-squares methods based on first-order partial differential equations because the least-squares formulation automatically employs the original first-order differential system as outflow boundary conditions, which are consistent. Previous computational experiments on the backward-facing step flow in a long channel show that calculation based on a least-squares method can be done accurately on a very small computational domain.

Parallel Algebraic Multigrid Solvers–Least Squares (AMS–LS) Applied to Hyperbolic Partial Differential Equations (PDEs)

Hans De Sterck
University of Colorado, Boulder
LLNL Contact
Robert Falgout

Summary

The objective of this project is to apply parallel algebraic multigrid (AMG) solvers to least-squares (LS) finite element (FE) discretizations of partial differential equations (PDEs) to obtain a solution method for these PDE problems that scales well on large parallel computers. In particular, we investigate the scalability of the AMG–LS approach for hyperbolic PDEs. Hyperbolic PDEs arise in many domains of science and engineering, including fluid dynamics and plasma dynamics, with applications including aerospace engineering, nuclear fusion, and astrophysics. In many of these applications, the problems in need of solution are extremely large. Large parallel computers are available, but present-day methods are not scalable to them.

The first part of the project was to couple the serial LS FE code FOSpack, which was developed at the University of Colorado (CU) at Boulder by John Ruge, to the parallel solver Library Hypre, which was developed at Lawrence Livermore National Laboratory's Center for Advanced Scientific Computing in a project directed by Rob Falgout. FOSpack, which is written in Fortran, has been coupled with the Hypre library, using Hypre's Fortran calling interface. The purpose of this coupling is threefold. First, the coupling provides an infrastructure with which to compare the FOSpack and Hypre AMG solvers. Second, it makes the other Hypre solvers available for use in FOSpack. Third, the coupling makes all the FOSpack features (for example, advanced equation parser, adaptive grid refinement, full MG, nonlinearity, higher-order elements) available for tests with the Hypre solvers. For the second part of the project, a preliminary parallel driver for Hypre has been written that implements the LS discretization of linear hyperbolic PDEs. It had been found before that an augmented equation system scales well with a serial AMG code, using a slightly modified AMG interpolation formula that takes into account the sign of the off-diagonal matrix elements. We have investigated parallel scaling of this formulation using the BoomerAMG solver provided in Hypre. The modified interpolation was implemented in BoomerAMG by Ulrike M. Yang.

It is the goal of the CU Boulder scalable solvers group to make the FOSpack code parallel. Coupled with Hypre, this will allow scientists and engineers to research parallel scalable solvers for a wide variety of LS-discretized PDE problems, including elasticity, fluid dynamics, and coupled systems, by using features such as adaptive grid refinement and higher-order elements. A parallel FOSpack–Hypre code will allow solving large complex science and engineering problems in a scalable way. For the particular application of hyperbolic PDEs, further large-scale parallel scaling tests will be performed using Hypre–BoomerAMG, and scalable AMG algorithms for hyperbolic PDEs will be further developed.

Performance Evaluation of High-Speed Interconnects Using Microbenchmarks and Scientific Applications

Rod Fatoohi
San Jose State University
LLNL Contact
Jeffrey Vetter

Summary

Our motivation is to study the affect of high-speed interconnects on the performance and scalability of scientific applications.

We are developing a set of microbenchmarks to characterize different aspects of the interconnects. Besides the traditional network parameters (latency and bandwidth), we employ parameters influenced by the network topology (such as bisection bandwidth and network diameter). Our microbenchmarks are implemented using two network APIs: the message passing interface (MPI) and the BSD sockets interface. We also employ several MPI-based scientific applications that use a variety of communication patterns. On this project, we are using the ASCI Purple benchmark codes.

Our microbenchmarks and applications have been tested on several systems using different interconnects, including two IBM SP switch-based systems (Frost and Blue) and two Quadrics-based clusters (pengra/MCR and TC2K). Future testing includes Cray Xl, an InfiniBand-based cluster, and BlueGene/L.

An Overlapping Schwarz Method for Nonlinear Unstructured Finite Element Elliptic Problems

Leszek Marcinkowski
Warsaw University
&
Xiao-Chuan Cai
University of Colorado, Boulder
LLNL Contacts
Panayot S. Vassilevski
&
Carol S. Woodward
CASC

Summary

We are interested in solving the nonlinear algebraic equations typically arising in the situation of nonlinear second-order elliptic PDEs discretized using finite elements. These problems are very difficult to solve as the nonlinearities involved are often unbalanced.

Few years ago Cai and Keyes proposed a new approach for solving nonlinear problems called Additive Schwarz Inexact Newton Method (ASPIN), which replaces the original nonlinear system of equations with a new one which generally has less difficult nonlinearities. ASPIN generates nonlinear subproblems that can be solved in parallel. The outputs of these subproblems are used to construct a global nonlinear system, which is solved by an outer Newton method. Therefore, Newton's method on small sparse problems in the original coordinates is used inside of a primary Newton loop on a transformed problem that possesses a (generally) dense Jacobian. The triumph is that the linear problems generated by the outer Newton method on the transformed problem can be preconditioned by sparse solves on subdomains of the original problem.

However for finely resolved problems on a large number of processors there is a need to introduce a coarse grid which would possibly reduce the number of linear iterations on each Newton step. This is challenging since there is no clearly best way to construct a coarse problem for the transformed system. Therefore we wanted to combine this approach with the concept of subspaces that are defined over the whole domain. The mesh is fine in every local subdomain and is gradually coarsened away from it. Such spaces were proposed first for linear problems by R.E. Bank and M. Holst. In the new setting we expected that we would not have to use additional coarse grid to reduce the number of linear iterations in Newton method.

The objective of the present paper is to construct a coarser version of the original fine grid nonlinear problem based on algebraic means targeting nonlinear problems discretized on generally unstructured grids. The coarsening is carried out outside a given domain G which is assumed to be a union of fine grid elements. The method is a generalization of the element agglomeration AMGe.

We confirmed that our algorithm is convergent with a rate bounded independently of the mesh size, i.e., the number of nonlinear Newton iterations and linear iterations remain constant independently of the mesh and they also remain fairly insensitive with respect to the number of subdomains. For our experiments we used 'hypre' package and the graph partitioner software 'METIS'. We also made progress in developing theory which will demonstrate that for a class of nonlinear problems our Newton method is locally convergent.

Automatic Discovery and Interaction with Bioinformatics Web

Anne H.H. Ngu,
Southwest Texas State University
&
Daniel Rocco
Georgia Institute of Technology
LLNL Contact
Terence Critchlow

Summary

The World Wide Web provides an incredible resource to genomics researchers in the form of query access to distributed data sources—e.g., BLAST sequence homology search interfaces. The number of these autonomous sources and their rate of change outpace the speed at which they can be manually classified, meaning that the available data is not being used to its full potential. Manually maintaining a wrapper library will not scale to accommodate the growth of genomics data sources on the Web, challenging us to produce an automated system that can find, classify, interact, and wrap new sources without tedious and low-level coding of wrappers. Previous research has not addressed the problem of automatically locating, classifying, and integrating classes of Bionifomatics Web data sources. A correct classification of any kind of Web data source must address both the domain of the source and the conversational/interaction semantics that is inherent in the design of many of the Web sources. In this paper, we proposed a novel approach to classification of Web data sources that takes into account both the capability and the conversational semantics of the source. The goal of my summer assignment is to enhance the current classifier at LLNL with the ability to discover the interaction pattern of a Web source that will lead to increased accuracy of the classification process. At the same time, it enables the extraction of process semantics that are necessary for the automatic generation of wrappers that can interact correctly with the sources.

I first identify the characteristics of the interaction pattern for BLAST data sources. Two algorithms are implemented for identifying the common interaction patterns. Three shortcomings are identified in the first heuristics-based algorithm. A second more robust and efficient algorithm based on computing page difference (PageDiff) between two HTML pages is proposed. A set of experiments is conducted that demonstrate that the PageDiff approach out performs heuristics by a factor of four. Integrating the PageDiff algorithm with the current LLNL classifier resulted in about a 10% increase in the number of correctly classified sources.

In the future, I want to pursue two goals. The first one is the identification of more complex interaction patterns that require human input. The second is the optimization of BLAST queries. Given that many BLAST sources could answer a particular query, which source should be selected at runtime? How is correlation between sites found out during the classification process? Can the quality of service be inferred during the classification process and used for finding out the correlation? If funding is available, I plan to follow up the collaboration with future visits to Livermore. I also plan to apply for a joint grant with the Data Science Group in the area of efficient and reliable access to Web sources.

Adaptive Mesh Refinement for Viscous Fingering in Porous Media

John Trangenstein
Duke University
LLNL Contact
Richard Hornung

Summary

During his sabbatical visit to LLNL, John Trangenstein worked on porting his adaptive mesh refinement code to several LLNL clusters, on extending the modeling, and on improving the algorithms used in this code.

Together with Andy Wissink and David Hysom of the Center for Applied Scientific Computing (CASC), Trangenstein adapted the Structured Adaptive Mesh Refinement Application Infrastructure (SAMRAI) BinaryTree, BoxTopology, and SpatialKey C++ classes to his code. These algorithms are useful in reducing the computational complexity of regridding, determining the communication between grid patches, and assigning patches to processors in order to reduce the need for communication.

Conversations with Robert Falgout and Panayot Vassilevski were useful in both debugging and improving the performance of multigrid algorithm in Trangenstein's porous media Adaptive Mesh Refinement (AMR) code. By incurring some additional communication costs in the multigrid smoother, he was able to reduce the number of multigrid iterations by about 25%. Trangenstein also completed the debugging of the 3D adaptive multigrid solver. These developments allowed simulation of 2D and 3D viscous fingering and beginning analysis of systematic errors in the Todd-Longstaff mixing model. Trangenstein also developed modifications of the Colella corner transport upwind scheme for linear advection with a pre-specified divergence-free velocity field. Jeffrey Hittinger was very helpful in getting this research started. The analysis of the 3D stability conditions has not yet been completed, so this scheme has not been implemented yet.

Trangenstein made some useful modifications to the Deferred Execution Tool developed by Bill Allard at Duke, replacing some synchronous communications with asynchronous ones and reducing the number of messages needed in communications between patches. These changes reduced the communication costs by about 25%. Finally, he modified the interactive graphics for adaptive mesh refinement to operate with distributed processor arrays. This was somewhat tricky because 3D graphics makes greater use of callbacks for features such as image rotation and cut-planes.

Duke graduate student Wenjun Ying accompanied Trangenstein during the sabbatical. Ying worked on hybrid adaptive mesh refinement code for modeling electrical wave propagation in the heart. His approach is to maintain an array of grid cells within a given level of refinement, rather than an array of grid patches. This gives him somewhat greater flexibility to deal with heart geometry. Space-filling curves will be used to subdivide the grid cell array into sub-arrays on separate processors. During this visit, Ying programmed and debugged a 3D adaptive multigrid algorithm for the elliptic and parabolic equations in the cardiac bidomain model. Incorporation of reactions will follow.

Continued

John Trangenstein
Duke University

Summary continued

Through previous employment at LLNL and association with a former Ph.D. student (Richard Hornung) and a former collaborator (Xabier Garaizar) now working at LLNL, Trangenstein had good ties to LLNL before his sabbatical visit. There are several research projects that could help to continue such connections. In order to make adaptive mesh refinement more efficient for computations involving wells, it would be helpful to use cylindrical grids around the wells and Cartesian grids between the wells. Such an approach would require irregular communications between grid patches; several LLNL CASC personnel have experience in this approach.

Compressible multi-phase flow in porous media involves nonlinear interactions between mass-conservation and the velocity/pressure field. The solution of these nonlinear equations is expected to completely dominate the computational work in black-oil models and would need to be implemented efficiently in adaptive mesh refinement. Communication with CASC member Peter Brown could be very helpful in this work.

CASC post-doctoral fellow Miguel Dumett, soon to become assistant professor at the University of Southern California, is interested in developing a collaboration with John Trangenstein in work on modeling electrical wave propagation in the heart.

Two–Grid Operators and Coarse Space Construction in Algebraic Multigrid Methods

Ludmil Zikatanov
&
Rolf Ryham
Pennsylvania State University
LLNL Contact
Rob Falgout

Summary

The main ingredients in constructing an algebraic multigrid methods construction are building a hierarchy of coarse spaces (coarsening) and applying an approximate solver in each of these spaces (smoothing). The development of an efficient and robust algorithm applicable to wide class of problems requires careful numerical and theoretical investigation of the relationships between approximations in different norms and local–global behavior of linear operators (obtained by various discretization techniques). In collaboration with Rob Falgout and Panayot Vassilevski from Livermore's Center for Advanced Scientific Computing, we have worked on the convergence analysis for two-grid methods aiming to apply the theoretical results in the design of more efficient algebraic multigrid methods. We have derived a sharp bound on the convergence for a general type of two-level product iterative methods (in fact, we have an identity for the norm of the error transfer operator for these methods, not just the bound). Such a result not only justifies the use of various quality measures used in the coarsening phase, but in most cases, it also provides necessary and sufficient conditions for uniform (say, with respect to mesh parameters) convergence.

We have also studied (and plan to continue to study) several coarsening algorithms based on this estimate. The idea is to use the smoother to find first a subspace, which is complementary to a coarse space, and then construct the coarse space by using a minimization of appropriate energy functional. This approach allows the use of a prescribed sparsity pattern of the interpolation matrix and an adaptive choice of coarse spaces as well. The possible applications of the techniques we have studied are the algebraic multigrid methods and some of the numerical homogenization methods. We plan to test these numerical solution algorithms by applying them to discretized systems of partial differential equations, such as linear elasticity and Maxwell equations, and to further study the possibility of building smoothers and coarse spaces iteratively, i.e., not only vary and adopt the hierarchy of coarse spaces but also appropriately change the smoother during multigrid iterations.

Institute for Scientific Computing Research

University Collaborative Research Program — Annual Research Reports

University Collaborative Research Program — Annual Research Reports

The University Relations Program (URP) at the Lawrence Livermore National Laboratory (LLNL) fosters collaboration between LLNL researchers and faculty from campuses of the University of California (UC) that have the potential for unique collaborations. Major objectives of the University Collaborative Research Program (UCRP) component of the URP are to encourage original work that has the potential to significantly impact research in areas of LLNL missions and to train future Laboratory employees and faculty members with specialization in these mission areas. The ISCR portion of the UCRP program provides support for graduate students (and sometimes short-term support for postdoctoral researchers) at any UC campus. It also provides an opportunity for graduate students to interact directly with Laboratory researchers through visits and joint activities. This section summarizes the progress on ISCR's UCRP Research Grants for FY 2003.

Scalable Algebraic Domain Decomposition Preconditioners

Randolph E. Bank
Principal Investigator
UC San Diego

Shaoying Lu,
Student,
UC San Diego

Charles H. Tong
Collaborator
LLNL

Panayot S. Vassilevski,
Collaborator
LLNL

Summary

Many leading edge scientific and engineering simulations expend large amounts of computational resources for the solution of linear systems of equations. Multilevel and domain decomposition methods have been identified as potentially scalable linear system solvers on terascale computer platforms. However, interprocessor communication overheads, degree of parallelism (in solving the coarse problems), and effects of increasing number of processors on convergence rates, all contribute to the list of obstacles to true scalability. The novel feature of our domain decomposition approach is that the subproblem residing in each processor is defined over the entire domain, although the vast majority of unknowns for each subproblem are associated with the subdomain owned by the corresponding processor. This feature ensures that a global coarse description of the problem is contained within each of the subproblems. The advantages of this approach are that interprocessor communications are minimized in the solution process while optimal order of convergence rates is preserved, and the speed of local subdomain solves can be maximized using the best existing sequential algebraic solvers.

This procedure is similar in philosophy to the parallel adaptive mesh refinement paradigm introduced by Bank and Holst, except that the present project deals with an algebraic version of the Bank-Holst paradigm in the sense that, instead of mesh refinement here, we coarsen the degrees of freedom (or matrix) outside the prescribed subdomain. Thus, the domain decomposition method applied in each processor involves the local subdomain plus a small coarse space defined on the whole domain outside the processor, and each solver utilizes only a single interprocessor communication to retrieve the global vector. This approach can be applied to general sparse matrices, although matrices arising from discretization of partial differential equations are the principal target.

Our parallel preconditioning algorithm for solving the sparse linear system $Ax = b$ is defined as follows:

Algorithm: FocusDD solver

$$f_p = \pi'_p * f, \text{ performed in parallel for } p = 0, 1, \cdots, P - 1$$

$$A_p = \pi'_p * A * \pi_p, \text{ performed in parallel for } p = 0, 1, \cdots, P - 1$$

$$\text{solve} \quad A_p * x_p = f_p, \text{ performed independently for } p = 0, 1, \cdots, P - 1$$

$$x = \text{collection of } \pi_p * x_p,$$

where P is the number of processors used. This algorithm uses a set of rectangular matrices $\{\pi_p\}$ (prolongation operators) and a set of coarse matrices $\{A_p\}$ (A_p and π_p are stored in processor p), which are to be constructed and stored in sparse parallel matrix format. f_p and x_p are the coarsened right hand side and coarsened solution, respectively, for processor p.

Continued

Randolph E. Bank
Principal Investigator
UC San Diego

Shaoying Lu,
Student,
UC San Diego

Charles H. Tong
Collaborator
LLNL

Panayot S. Vassilevski,
Collaborator
LLNL

Summary continued

For the year of 2003, we focused on fine-tuning the parallel algebraic multigrid (AMG) solvers, implemented in the software package FocusDD. In addition to our earlier two-level solvers, we have developed several new three-level solvers. The best performing among them is constructed as an additive Schwartz method, with each major iteration composed of both three-level and two-level solution steps. This FocusDD solver has been used as a preconditioner for the Generalized Minimal Residual (GMRES) method. The scalability of this solver has been studied, in comparison to that of the BoomerAMG solver, on the 16-dual-processor Beowulf cluster in the Scientific Computing Group at UC San Diego

We are preparing a report entitled *Scalable Parallel Algebraic Multilevel Methods* describing this work. Shaoying Lu will present a lecture on this work at the SIAM Conference on Parallel Processing for Scientific Computing, to take place in San Francisco on February 25–27, 2004.

We conclude with a short example. The 3D Poisson equation is solved with the FocusDD solver. A scalability study is conducted with varying number of processors (np) and a fixed number of unknowns per processor (npp). With np ranges from 2 to 128 and npp = 1000, the convergence rate and the computing time of our FocusDD solver is compared with those of the BoomerAMG solver in HYPRE. As shown in table 1, the FocusDD solver is less efficient than BoomerAMG with less than 64 processors, in both initialization and solution stages. However, the solution time of both solvers becomes comparable for 64 processors. Furthermore, the FocusDD solver becomes more efficient than the BoomerAMG with 128 processors. This trend indicates that the FocusDD solver may have better parallel efficiency than the BoomerAMG solver in the solving stage, for large clusters.

np	FocusDD				BoomerAMG			
	iter	γ	setup time	solve time	iter	γ	setup time	solve time
2	4	0.026889	0.85	0.09	1	0.0	0.07	0.05
4	4	0.028566	1.33	0.10	1	0.0	0.12	0.10
8	7	0.098530	8.47	0.53	1	0.0	0.42	0.25
16	7	0.089256	14.90	0.82	1	0.0	1.56	0.60
32	8	0.164202	17.73	1.03	1	0.0	3.65	1.08
64	8	0.161620	89.40	3.43	1	0.0	16.45	4.83
128	8	0.167599	161.51	3.85	1	0.0	159.66	17.55

Table 1: Convergence behavior of the FocusDD solver

Computationally Efficient Example-Based Image Segmentation

Serge Belongie
Principal Investigator
UC San Diego

Josh Wills
Student,
UC San Diego

Sameer Agarwal
Student
UC San Diego

Imola K. Fodor
Collaborator
LLNL

Summary

The proliferation of digital cameras, both still and video, has produced massive amounts of visual data everywhere—from video surveillance stations to hard disks from personal low-cost digital cameras. Effective tools for image segmentation offer the potential to intelligently organize and query such collections.

Image segmentation is the problem of partitioning the pixels in an image into a relatively small number of regions that correspond to objects or parts of objects. It is one of the hardest (and oldest) open problems in computer vision and plays an important role in the process of object detection and recognition. As challenging and computationally intensive as image segmentation is, it also happens to be a problem that the human visual system solves effortlessly.

The goal of this project is to develop methods for image and video segmentation with an emphasis on motion-based processing. In particular, our two main interests are (a) image sequences containing objects undergoing large inter-frame motion (e.g., more than 10% of the image width) and (b) footage of objects in motion exhibiting temporal periodicity (e.g., a pedestrian or a running dog). Examples of practical applications likely to benefit from the proposed research include:

- tracking people, vehicles, and animals in surveillance video.
- developing vision systems for the visually impaired (e.g., a digital "seeing eye dog").
- locating television shows based on content (e.g., "find me a nature show on cheetahs").

Since beginning work on the project, we have developed several new algorithms that have led to promising results in a variety of problem domains involving segmentation and/or clustering. We also produced a number of publications that have been accepted or are under review.

The paper "Segmentation by Example" [1] follows directly from the proposed research activity. In this work, we operationalized the idea of inheriting local graph connections from human-labeled examples of segmented images and applied it to real images. We currently have it working on static images and are extending it to video sequences. Over the last year, several works have appeared from other research groups on learning-based approaches to image segmentation, and there appears to be a great deal of interest in this area among computer vision researchers.

Our explorations thus far in video segmentation have led to two papers — "What Went Where" [4] and "A Feature-based Approach for Determining Dense Long Range Correspondences" [5]—that extend the layer-based segmentation framework to sequences with inter-frame motion far beyond the capability of any previously existing

Continued

Serge Belongie
Principal Investigator
UC San Diego

Josh Wills
Student,
UC San Diego

Sameer Agarwal
Student
UC San Diego

Imola K. Fodor
Collaborator
LLNL

Summary continued

algorithm. Josh Wills gave the talk for "What Went Where" at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in June 2003. The second paper is currently in review for the 2004 European Conference on Computer Vision (ECCV2004). One application area for this line of work is the deletion of unwanted objects in a series of photographs, e.g., in virtual set design for special effects.

Finally, a clustering algorithm (due to Hochbaum and Shmoys) that we encountered while pursuing the "segmentation by example" framework turned out to be ideally suited to a problem in computer graphics that was brought to our attention by Prof. Henrik Wann Jensen of UC San Diego. We pursued the application of this algorithm to the problem of accelerating image-based rendering using environment maps and the resulting paper [2] was accepted for a talk delivered by Sameer Agarwal at the 2003 meeting of the Special Interest Group in Computer Graphics (SIGGRAPH).

We are pursuing extensions of the above approaches in motion segmentation to the problem of tracking approximately temporally periodic objects in surveillance footage (e.g., pedestrians, a running dog, a bird flapping its wings). We submitted a paper called "Structure from Periodic Motion" [3] with preliminary results in this area to ECCV 2004.

Publications

1. Sameer Agarwal and Serge Belongie, *Segmentation by example*, UC San Diego, Technical Report CS2003-0762 (2003).

2. Sameer Agarwal, Ravi Ramamoorthi, Serge Belongie, and Henrik Wann Jensen, *Structured importance sampling of environment maps*, in SIGGRAPH, San Diego, CA (2003).

3. Serge Belongie and Josh Wills, *Structure from periodic motion*, UC San Diego, Technical Report CS2003-0767 (2003).

4. Josh Wills, Sameer Agarwal, and Serge Belongie. "What went where," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, volume 1, (2003), pages 37–44

5. Josh Wills and Serge Belongie, *A feature-based approach for determining dense long range correspondences*, UC San Diego, Technical Report CS2003-0768 (2003).

Fast Direct Solvers for Large, Dense-Structured Matrices

Ming Gu
Co-Principal Investigator
UC Berkeley

S. Chandrasekaran
Co-Principal Investigator
UC Santa Barbara

Timothy Pals
Student
UC Santa Barbara

Daniel White
Collaborator
LLNL

Summary

Our award is primarily concerned with the design of fast direct solvers for large, dense-structured matrices. Classical examples of large, dense-structured matrices include Toeplitz and discrete Fourier matrices. However, the kinds of matrices we are concerned with typically arise in the solution of partial differential equations and their associated integral equations. In particular they arise from the direct discretization of integral equations (including those of scattering theory) and also as the Schur complements of finite element or finite difference discretizations of partial differential equations. These are large dense matrices with complicated structures. The challenge has been to harness these structures efficiently to construct fast solvers and effective preconditioners.

Indeed a great number of algorithms have been developed to exploit such matrix structures over the last 15 years, starting with the Fast Multipole Method (FMM) of Rokhlin and his colleagues. The main innovation was to replace large subblocks of the dense matrices by low-rank approximations to significantly speed up matrix-vector multiplications. This enabled the use of iterative techniques like the conjugate-gradient method and the Generalized Minimal Residual (GMRES) method to solve the dense system of equations rapidly.

However, the speed of convergence of iterative methods is problem dependent. In fact, iterative methods can be very slow to converge or may not converge without the availability of a good preconditioner. For example, for near resonant and strong multiple scattering problems, a large number of iterations are usually required for convergence. There is now a whole area of research that looks into whether many existing integral equations can be rewritten in forms that require much fewer iterations for iterative methods. So far, some limited success has been achieved.

To get some appreciation of the complexity of the matrix structure involved, we consider the matrix whose (i,j)th entry is $\log|x_i-x_j|$, where the x_i are points that are evenly (homogeneously) distributed between 0 and 1. Each off-diagonal block of this matrix can be approximated by a submatrix whose rank does not exceed some small number depending on the precision (< 10 for single precision). Other partitions of this matrix are also possible depending upon the expansions used. However, the structure becomes even more complicated in 2D and 3D. Even points distributed homogeneously on a curve in two-dimensional space can lead to a complex structure if the curve has a complex geometry.

We showed how a ULV^T factorization could be used to design fast backward stable algorithms for this class of problems in one dimension. Our success is based on three novel ideas: First, we developed a new algebraic matrix structure called a sequentially semi-separable algebraic structure; basically, this is the matrix structure of the inverse

Continued

Ming Gu
Co-Principal Investigator
UC Berkeley

S. Chandrasekaran
Co-Principal Investigator
UC Santa Barbara

Timothy Pals
Student
UC Santa Barbara

Daniel White
Collaborator
LLNL

Summary continued

of semi-separable matrices and versions of it were discussed by Dewilde and van der Veen in the context of time-varying linear systems. Then we showed that linear systems of equations involving such matrices could be solved in $O(n\,r^2)$ flops, where n is the matrix dimension and r is the rank of off-diagonal blocks. Finally, we showed that the coefficient matrices arising from discretization of certain integral equations could be compressed into this form with an efficient compression algorithm. While this algebraic structure is specifically designed for problems in one spatial dimension, it can be applied to any matrix, even those that arise from problems with three spatial dimensions.

We showed that the peak off-diagonal ranks are quite low for matrices that arise from the global spectral discretization of the two-dimensional exterior scattering problem using the method of Kress. This method is of particular interest to our LLNL collaborators and their EIGER project. Again, this is a very high-order spectral method, and there was no fast matrix-vector multiplication algorithm even for the associated matrix. Again, as can be seen our one-dimensional algebraic approach captures the structure very efficiently and produces a fast direct solver too. This was a very pleasant surprise. Part of our proposed work is to work out a theoretical reason for the surprising effectiveness of the sequentially semi-separable algebraic structure.

We have implemented and tested the algorithm in several cases and have had good results with timing and accuracy [1,2,6]. Our numerical experiments show that our algorithm is a couple of orders of magnitude faster than standard methods, for practically important ranges of the peak off-diagonal rank.

Many of the examples we tested include those provided by our LLNL collaborators. Our numerical experiments show that the sequentially semi-separable structure is even doing a good job on problems with three spatial dimensions.

We have investigated ULV^T factorization based fast algorithms for two-dimensional structures [5]. In particular we have produced an algebraic characterization of such matrices, and a general fast solver. The algebraic characterization is obtained by viewing the sequentially semi-separable structure as being defined on a unary tree (every node has at most one child).

As is evident we have emphasized the production of codes that can be used to estimate the actual run-times of all our algorithms. That is, we are not implementing our algorithms in Matlab. Obviously the downside of this approach is that as we start working with higher dimensional problems it takes a significant amount of work to produce good codes. Our LLNL collaborators, who have been providing us with example matrices, have helped us tremendously.

Continued

Ming Gu
Co-Principal Investigator
UC Berkeley

S. Chandrasekaran
Co-Principal Investigator
UC Santa Barbara

Timothy Pals
Student
UC Santa Barbara

Daniel White
Collaborator
LLNL

Summary continued

The thesis work of Timothy Pals, a graduate student being supported on this project, has already produced some outstanding results. Pals developed the first high-order technique for solving two-dimensional scattering problems that is amenable to acceleration by the fast multipole method. He also developed the first stable version of the fast multipole method for the two-dimensional scattering problem and invented some fast direct solvers that are directly tuned for two-dimensional scattering problems.

Publications

1. S. Chandrasekaran and M. Gu, *Fast and stable algorithms for banded plus semi-separable matrices*, submitted to SIAM J. Matrix Anal. Appl. (2000).

2. S. Chandrasekaran and M. Gu, *A fast and stable solver for recursively semi-separable systems of equations*, in *Structured matrices in mathematics, computer science and engineering, II*, edited by Vadim Olshevsky, in the Contemporary Mathematics series, AMS publications (2001).

3. S. Chandrasekaran and M. Gu, "Fast and Stable Eigendecomposition of Symmetric Banded plus Semi-separable Matrices," 1999, *Linear Algebra and its Applications*, Volume 313, Issues 1-3, 1 (July 2000), pages 107–114.

4. S. Chandrasekaran and M. Gu, *A Divide-and-Conquer Algorithm for the Eigendecomposition of Symmetric Block-Diagonal Plus Semiseparable Matrices*, 1999, accepted for publication in *Numerische Mathematik*.

5. S. Chandrasekaran and M. Gu, *Superfast Nested Dissection* (in preparation 2004).

6. S. Chandrasekaran, P. Dewilde, M. Gu, T. Pals, and A.-J. van der Veen, *Fast Stable Solvers for Sequentially Semi-separable Linear Systems of Equations*. Submitted to SIMAX (2002).

Numerical Study of Coexisting Superconductivity and Ferromagnetism

W. E. Pickett
Principal Investigator
UC Davis

Alan Kyker
Student
UC Davis

F. Gygi
Collaborator
LLNL

Summary

Superconductivity and ferromagnetism are the two most evident macroscopic manifestations of quantum mechanical behavior, but these macroscopic quantum states are strongly antagonistic. On several grounds, it has been expected that they could never coexist; yet recently three examples have been identified (UGe_2, $ZrZn_2$, and URhGe). The scientific questions are several and basic: how does the supercurrent cope with the (frozen in) magnetic flux, which it normally abhors; how does the superconducting pairing arise; what is the character of the low-energy excitations? Almost 50 years ago Vitaly Ginzburg, who shared the 2003 Nobel Prize in Physics, concluded that electrodynamic considerations precluded coexistence of ferromagnetism and superconductivity. His considerations did not take into account the possibility of a "spontaneous flux phase," which is the arrangement that allows the supercurrent to tolerate the magnetic flux.

During the initial year of this grant, it has been established that—at the Ginzburg-Landau level (the same Nobel laureate Ginzburg) involving minimization of the free energy of the system near the superconducting critical temperature—an intrinsic magnetic flux is equivalent to an externally applied field. Minimization of the free energy is accomplished by solving second-order n on-linear differential equations for the complex superconducting order parameter $\vec{A}(\vec{r})$ and the electromagnetic vector potential $\psi(\vec{r})$, for which a steepest descent algorithm is used.

A second basic question concerns how the spin degree of freedom of the constituents of the pair is affected by the magnetization and involves the superconducting gap equation itself. The Fulde-Farrell-Larkin-Ovchinnikov (FFLO) solution arose from treatment of the gap equation in the presence of (magnetic) exchange splitting, which leads to new solutions that correspond to an inhomogeneous superconducting order parameter. This part of the project is involved with determining how the electron-dispersion relation (energy vs. momentum) impacts the formation of the FFLO phase.

Each of the simulations can he controlled by a graphical user interface (GUI), which graduate student Alan Kyker put together using C code and the OpenGL graphics library. A high quality and versatile visualization capability is *essential* for this project, because there is simply no alternative to direct visualization for understanding the character of the solution. The visualization code has two very useful features (besides the obvious): (1) it refreshes the image periodically (user selectable from 1/60 sec) while the equations are being iteratively solved, so the progress of the algorithm can be assessed in real-time; (2) the program provides several 3D graphical real-time-selectable views of various aspects of the simulation including magnetic field $\vec{B}$ field and magnitude/phase of the complex order parameter, with orientations controlled by the mouse.

Continued

W. E. Pickett
Principal Investigator
UC Davis

Alan Kyker
Student
UC Davis

F. Gygi
Collaborator
LLNL

Summary continued

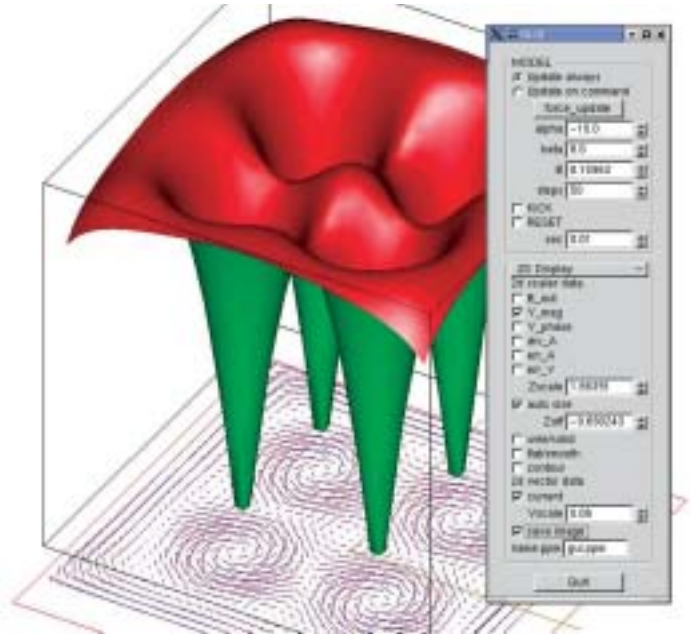


Fig. 1. Surface plot of the magnitude of the (complex) superconducting order parameter for a superconducting square wire from the code *vortex*, showing four magnetic vortices that are forced upon the system by the magnetism. Bottom: circulation of the superconducting current around each vortex. Superimposed at right is the GUI that allows easy control of parameters and re-execution (in actual applications, it does not overlay the surface plot).

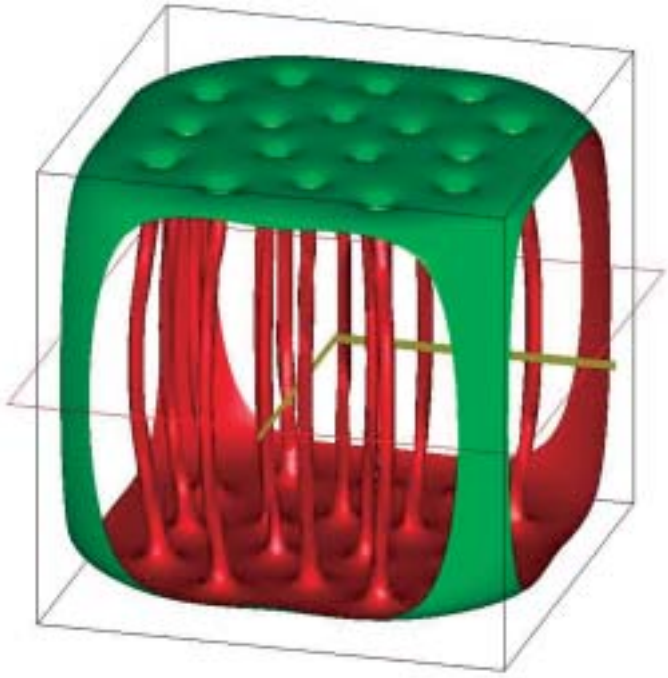


Fig. 2. Isosurface plot of the magnitude of the superconducting order parameter for a superconducting ferromagnetic cube. The spontaneous vortex lattice is evident, approaching the hexagonal arrangement predicted long ago for conventional superconductors by Alexei Abrikosov, co-recipient of the 2003 Nobel Prize in Physics.

Simulation Of Compressible Reacting Flows

S. Sarkar
Principal Investigator
UC San Diego

A. W. Cook
Collaborator
LLNL

Summary

A fundamental problem of interest regarding design of target capsules for the National Ignition Facility (NIF) is the behavior of a burn-front when it propagates through a DT mix contaminated by inert shell material introduced by Rayleigh-Taylor instabilities. In order to accurately describe burn-front propagation, we have developed a scheme that is high-order on arbitrarily non-uniform grids. Large computational grids are required to resolve the turbulence and burn chemistry. Long-time evolution is required to achieve converged statistics. Therefore, the simulations, run on the IBM Power4 machine, have stringent computational requirements. Each case is run on 128 CPUs and requires approximately 3,400 CPU hours.

The unsteady, three-dimensional form of the compressible equations for a reactive mixture of fluids is solved. A non-uniform grid is used in the x-axis, the direction of flame propagation, with clustering of points in the burn region. Derivatives in the x -direction are computed using the 6th-order compact scheme valid for arbitrary non-uniform grids. The code is parallelized using message-passage interface (MPI) and runs on the unclassified IBM SP machine at LLNL. We use a large computational box of size $680\delta_F \times 340\ \delta_F$ with δ_F denoting the flame thickness and a grid with 2304×1536 grid points. The integral length scale of the fluctuations is $l/\delta_F \approx 50$. Approximately 400,000 iterations are required to generate a time record that is sufficiently long for converged statistics.

In the zones contaminated by inert material, the temperature rise due to the burn-energy release is smaller than that in a pure DT mix leading to a lower reaction rate and a lower local-burn velocity. A mechanism for distorting the burn-front that is operative even in an uncontaminated mix is the so-called hydrodynamic instability, also called the Darrieus-Landau (DL) instability, that causes large-scale undulations of the front. Finally, Rayleigh-Taylor instabilities and associated turbulence in the DT mix cause wrinkling of the front.

At any given time, a burn velocity, $U_F\,(t)$, based on the overall reactant consumption rate can be defined. A time-average burn velocity, $\bar{U}_F$, can also be defined to characterize the long-time behavior of the burn-front. We have performed simulations with different levels of composition fluctuations,

$$Y' = 0,\ 0.02,\ 0.06,\ 0.18,$$

where Y' is the root-mean-square (rms) fluctuation in fuel-mass fraction. Velocity fluctuation levels measured by the rms velocity, u', have been varied,

$$u'/S_L = 0,\ 0.45,\ 1.1,\ 2.5,$$

corresponding to low-moderate turbulence levels. Here S_L is the burn velocity for laminar, one-dimensional propagation.

Continued

S. Sarkar
Principal Investigator
UC San Diego

A. W. Cook
Collaborator
LLNL

Summary continued

Figure 1 shows the time history of the burn velocity in a mix with *zero* imposed velocity fluctuations. For a pure reactant mix, the evolution is smooth and the time-average burn velocity, $\bar{U}_F \approx 1.3$. The natural DL instability is responsible for the 30% increase of burn-front area and associated burn velocity. Figure 1 shows that, with increasing compositional fluctuations, the burn propagation becomes progressively unsteady and the time-averaged velocity also increases. Visualizations (not shown) suggest that the pockets of inhomogeneity seed strong DL instabilities that increase the burn-front area. Apparently, this effect dominates the competing trend of a decrease of local-burn velocity with increasing compositional inhomogeneity so that there is a net *increase* of burn velocity with increasing compositional fluctuation.

In general, one would expect both compositional and velocity fluctuations due to RT instabilities in the reactant mix. We have varied the compositional fluctuation level for different choices of turbulence levels. Figure 2 compares the behavior of the time-averaged burn velocity, $\bar{U}_F$, at a low turbulence level, $u'/S_L = 0.45$, with that at zero-velocity fluctuation. A number of conclusions can be drawn. First, for the pure reactant mix ($Z' = 0$), the burn velocity significantly increases when there are velocity fluctuations in the mix. Second, although mixture inhomogeneity, i.e., nonzero Z', generally tends to increase the burn velocity, its influence is weaker when velocity fluctuations are simultaneously present in the reactant mix. Third, at high levels of mixture inhomogeneity, the burn velocity appears to be relatively insensitive to the level of velocity fluctuations.

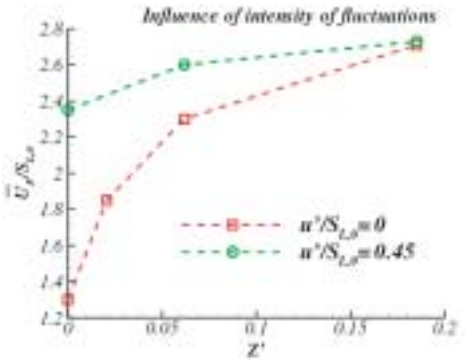


Fig. 1 The time evolution of the burn velocity.

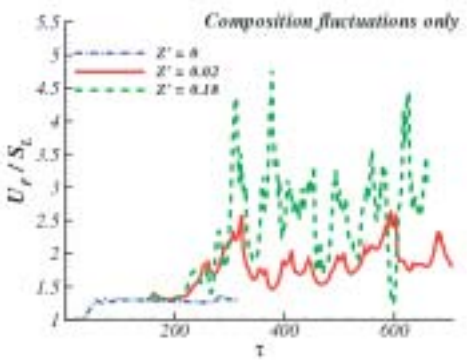


Fig. 2 The time-averaged burn velocity for different cases.

Publications

D. G. Lopez and S. Sarkar, "Effect of imperfect mixing on flame propagation,"*Bulletin of the 56th Annual meeting of the American Physical Society*, Division of Fluid Dynamics, New Jersey, 48 (10), 169 (2003).

Probabilistic Clustering of Dynamic Trajectories for Scientific Data Mining

Padhraic Smyth
Principal Investigator
UC Irvine

S. Gaffney
Student
UC Irvine

Chandrika Kamath
Collaborator
LLNL

Summary

In this research project we have developed a set of flexible methods and algorithms for tracking and clustering time-trajectories of coherent structures (such as cyclones) in spatio-temporal grid data. These algorithms and software provide a basic set of data- analysis tools for exploration and modeling of dynamic objects, in a manner analogous to the much more widely available techniques for clustering of multivariate vector data such as k-means, Gaussian mixtures, hierarchical clustering, and so forth.

Traditional clustering algorithms typically assume that the N data objects to be clustered either exist as fixed-dimensional vectors or that a matrix of N^2 distances or similarities is available. When applied to trajectory data (e.g., latitude-longitude positions of a storm center as a function of time) both approaches have disadvantages. The first approach requires that all trajectories be converted to fixed-dimensional vectors of the same length; this is often an unnatural representation of the data, and the underlying spatial and temporal information is no longer explicitly represented. The second approach is also problematic in that a natural distance measure may be difficult to construct; furthermore, the $O(N^2)$ computational complexity required to construct such a matrix may be intractable for large values of N.

To address these problems we have investigated a probabilistic alternative to trajectory clustering that uses finite mixture models. This is based on a generative model for the trajectory data in the form of a mixture of regression "shapes." Under this model, each trajectory is assumed to be generated by one of K mixture components, and each component represents a mean shape (modeled as a polynomial or spline function) to which noise is added to generate an observed trajectory. The clustering problem (from a statistical viewpoint) is to recover both the parameters describing the K underlying shapes (regression coefficients) and a probability distribution over shapes (or clusters) for each trajectory. We developed a set of unsupervised learning algorithms based on the Expectation-Maximization (EM) procedure that can solve this clustering problem by maximizing the likelihood of the data (multiplied by an appropriate prior). We found that this technique for clustering curves tends to systematically produce models with better out-of-sample predictive power than models that vectorize the data (Gaffney, Robertson, and Smyth, 2001).

In the methodology described above, each trajectory within a cluster shares the same regression coefficients; they only differ from each other (from the model's viewpoint) in terms of the additive noise added to each. In Gaffney and Smyth (2003) we generalized this technique such that trajectories within a cluster are allowed some random deviations (in parameter space) from the overall mean shape for that cluster. The resulting learning problem is also based on EM, and in effect it is clustering the observed trajectories in

Continued

Padhraic Smyth
Principal Investigator
UC Irvine

S. Gaffney
Student
UC Irvine

Chandrika Kamath
Collaborator
LLNL

Summary continued

parameter space, while simultaneously estimating the parameters for each curve. In experiments on simulated curves and tracked cyclone trajectories we found that this method tended to produce models with better out-of-sample predictive power than the earlier ``fixed coefficient" method (Gaffney and Smyth 2003; Gaffney, 2004).

In another extension of this work, we developed techniques to handle the real-world problem of trajectories that are mismatched in time. For example, a tracking algorithm might miss the early part of a cyclone's "life" relative to a different cyclone that is picked up by the tracking algorithm very early in its lifecycle. The resulting latitude-longitude trajectories will not be aligned, making it less likely that trajectories will be clustered appropriately. Traditional approaches involve relatively simple global alignment followed by clustering. In Chudova, Gaffney, and Smyth (2003) we developed a new technique that allows each trajectory to have an unobserved "time-offset" parameter relative to the canonical life cycle for its cluster. We showed that the EM procedure can be used to derive a statistical learning algorithm that simultaneously estimates (a) the trajectory shape for each cluster, (b) a probability distribution over clusters for each trajectory, and (c) an estimate of the most likely time-offset for each trajectory. The resulting model again shows consistently better out-of-sample performance in systematic experiments than the earlier methods.

We applied our methodology to the problem of tracking and clustering of extra-tropical cyclones (ETCs) in the North Atlantic. Understanding ETC trajectories is scientifically important for understanding both the short-term dynamics and long-term variation of atmospheric processes, such as identifying how ETC frequency and intensity distributions may be related to global climate change.

In collaboration with Dr. Andrew Robertson (IRI [International Research Institute for Climate Prediction], Columbia) and Professor Michael Ghil (UCLA) we used data from the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM3) general circulation model (GCM), run with observed sea-surface temperatures specified at the lower boundary over the 1980–1995 period. For the tracking, we used the atmospheric pressure at mean sea level (MSLP) given on an approximate 2.8°× 2.8° Gaussian grid over the globe. The winds at 200 hPa (about 10-km elevation) were also available for use in interpreting the resulting cyclone clusters. The data are available every 6 hours and we analyzed data for the winter months (1 November to 30 April) from 1980 to 1995. We used a simple tracking algorithm to detect 614 cyclones of different durations, each with a minimum of 10 observations (i.e., at least 2.5 days long). From the resulting set of ETC trajectories we demonstrated that our clustering methodology was able to reveal robust and physically meaningful clusters of ETC patterns and behavior (Gaffney, Robertson, and Smyth, 2001; Gaffney et

Continued

Padhraic Smyth
Principal Investigator
UC Irvine

S. Gaffney
Student
UC Irvine

Chandrika Kamath
Collaborator
LLNL

Summary continued

al. 2004). For example, we found that cyclones tended to group into vertically, diagonally, and horizontally oriented paths in the Atlantic, with the horizontal-cluster (for example) consisting of cyclones that move horizontally (west to east) across the coastline of Europe.

Publications

S. Gaffney, A. Robertson, and P. Smyth, "Clustering of extra-tropical cyclone trajectories using mixtures of regression models," in *Proceedings of the Fourth Workshop on Mining Scientific Data Sets, Seventh* ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 15–20, August 2001.

S. Gaffney and P. Smyth, "Curve clustering with random effects mixtures," in *Proceedings of the Ninth International Workshop on AI and Statistics,* January 2003.

D. Chudova, S. Gaffney and P. Smyth, "Probabilistic models for joint clustering and time-warping of multidimensional curves" in *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers, pp. 134–141, August 2003.

S. Gaffney, A. Robertson, M. Ghil, and P. Smyth, "Probabilistic clustering of extra-tropical cyclones using regression mixture models," 2004 (in preparation).

S. Gaffney, Mixture Models for Clustering Sets of Curves, Ph.D. Thesis, University of California, Irvine, 2004 (expected).

Memory Access Pattern Signatures and Certificates of Relevance for Benchmarks

Allan E. Snavely
Principal Investigator
UC San Diego

Michael O. McCracken
Student
UC San Diego

Jeffery Vetter
Collaborator
LLNL

Bronis de Supinski
Collaborator
LLNL

ISCR Institute for Scientific Computing Research

Summary

The goal of the Department of Energy's (DOE's) Scientific Discovery through Advanced Computing (SciDAC) Program's Performance Evaluation Research Center (PERC) [see http://perc.nersc.gov/main.htm] is to develop a science for understanding performance of scientific applications on high-end computer systems and develop engineering strategies for improving performance on these systems. This project is integrating several active efforts in the high-performance computing community and is forging alliances with application scientists working on DOE's Office of Science missions to ensure that techniques and tools being developed are truly useful to end users.

The Performance Modeling and Characterization (PMaC) laboratory at the San Diego Supercomputer Center (SDSC) is a participant in PERC. It is working to bring scientific rigor to the prediction of scientific application performance on current and projected high-performance-computing (HPC) platforms with a special emphasis on the performance implication of the memory hierarchy. PMaC's goal is to predict the performance of applications more accurately than traditional benchmarking methods or traditional cycle-accurate performance simulations.

This research has investigated the performance implications of memory-access patterns and useful definitions of "signature distance" between the memory-access patterns of different basic blocks from the same or different programs. The goal is to improve the accuracy and speed of the Convolution method for performance prediction. We have enhanced the functionality of the MetaSim tool for gathering memory-access-pattern signatures and have made this tool platform-independent. We have been investigating what kinds of memory-access patterns exist "in nature" and exploring the performance implications of memory-access patterns. We have developed a nomenclature and symbolic representation of memory-access patterns, leveraging previous work by Nick Mitchell.

Michael McCracken has been examining definitions of "signature distance" between basic blocks—the idea being that (possibly) basic blocks that look roughly the same in terms of memory-access patterns may perform roughly the same on a given machine. In work defining a meaningful metric for "signature distance" he has established orthogonal properties of loops including memory footprint, memory-access pattern, type and intensity of floating-point, and Instruction-level parallelism (ILP) operations that could provide "certificates of relevance" for benchmarks. In other words, he has shown early evidence that it is possible to reasonably estimate the likely performance of an application that has been profiled by MetaSim if all of the application's basic blocks are mapped to a usefully similar set of benchmark basic blocks whose performance has already been established. McCracken has published a paper [1] showing how this approach may be applied for dynamic algorithm selection guided by performance model predictions. Currently he is applying the technique to a commercial code.

Continued

Allan E. Snavely
Principal Investigator
UC San Diego

Michael O. McCracken
Student
UC San Diego

Jeffery Vetter
Collaborator
LLNL

Bronis de Supinski
Collaborator
LLNL

Summary continued

McCracken studied about 108 scientific loops taken from a paper in *Parallel Computing* titled *A comparative study of automatic vectorizing compilers* by D. Levine, David Callahan, and Jack Dongarra. Using the statistical analysis tool SAS he was able to derive parameterized functions to predict the performance of these loops as a function of their memory footprint, memory-access pattern, floating-point mix, and intensity and ILP on IBM Power3 and Power4 processors. A poster is in preparation for the UC San Diego Research Review. Another paper is in preparation that will apply the modeling method to explain and predict the performance the commercial Computational Fluid Dynamics (CFD) application Cobalt (see http://www.cobaltcfd.com/).

Publications:

1. Michael O. McCracken, Allan Snavely, and Allen Malony, *Performance Modeling for Dynamic Algorithm Selection*, ICCS, Workshop on Performance Modeling and Analysis (PMA03), Melbourne (2003).

Statistical Inference from Microarray Data with Applications in Breast Cancer Research

Mark van der Laan
Principal Investigator
UC Berkeley

Annette Molinaro
Student
UC Berkeley

Dan Moore
Collaborator
LLNL

Summary

The advent of microarray technologies has made feasible the measurement of gene copy number and gene expression in very large numbers from a wide variety of biologic samples. It is believed that gene copy number and gene expression are important indicators of cancer progression and that an understanding of how genes are turned "on" or "off" will ultimately lead to better strategies for treatment and to a reduction in cancer incidence.

The overall goal of this collaboration is to link gene microarrays to disease progression. This goal includes the following components: to determine reliable methods for classifying tumors based on their genetic profile, to link the classified tumors to covariates and clinical outcomes, to link the gene microarrays to covariates, and to link these gene microarrays and covariates to clinical outcomes.

This year we initially focused on a data set consisting of 152 breast cancer cases, for which we had comparative genomic hybridization (CGH), pathological, and clinical covariates. Our goal was to find chromosomal loci where specific alterations would be predictive of survival time. In this project we proposed a unified strategy for estimator construction, selection, and performance assessment in the presence of censoring. This approach is entirely driven by the choice of a loss function for the full (uncensored) data structure and can be stated in terms of the following three main steps: (1) Define the parameter of interest as the minimizer of the expected loss, or risk, for a full data-loss function chosen to represent the desired measure of performance. Map the full data-loss function into an observed (censored) data-loss function having the same expected value and leading to an efficient estimator of this risk; (2) Construct candidate estimators based on the loss function for the observed data; and (3) Apply cross-validation to estimate risk based on the observed data-loss function and to select an optimal estimator among the candidates. A number of common estimation procedures follow this approach in the full-data situation, but depart from it when faced with the obstacle of evaluating the loss function for censored observations.

Tree-based methods, where the candidate estimators in Step 2 are generated by recursive binary partitioning of a suitably defined covariate space, provide a striking example of the chasm between estimation procedures for full data and censored data (e.g., regression trees as in Classification and Regression Trees [CART] for uncensored data and adaptations to censored data). Common approaches for regression trees bypass the risk-estimation problem for censored outcomes by altering the node-splitting and tree-pruning criteria in manners that are specific to right-censored data. In this project we implemented a generalization of regression trees to censored data. We used the CART (Breiman, et al. 1984) algorithm altered by the general-estimating equation methodology of Mark van der Laan in collaboration with James Robins (van der Laan & Robins 2003).

Continued

Mark van der Laan
Principal Investigator
UC Berkeley

Annette Molinaro
Student
UC Berkeley

Dan Moore
Collaborator
LLNL

Summary continued

The formalization of this approach encompasses univariate and multivariate prediction and density estimation for censored and non-censored data. This project has resulted in a technical report (www.bepress.com/ucbbiostat/paper135/) and a manuscript in press for a special issue, "Genomics," of the Journal of Multivariate Analysis.

Upon completion of this project, we began investigating other piecewise constant regression methods based on biological reasoning for generating candidate estimators in Step 2. CART evaluates and chooses the best binary splits of the covariates, building a list of *"and"* statements that predict survival times (e.g. loss at locus 1 *"and"* gain at locus 2 predicts survival of x months). However, due to reasoning that multiple chromosomal aberrations could lead to a similar or identical effect, CGH and expression data may require alternatives to the *"and"* statement ordering (e.g. loss at locus 1 *"or"* loss at locus 3 *"and"* gain at locus 2 predicts survival of y months). Our current project involves building a new algorithm which includes *"and"* and *"or"* statements. This algorithm is more aggressive than CART by allowing a variety of covariate splits and subsequently unions of those splits. Annette Molinaro is in the final stages of implementing this algorithm for univariate prediction in the statistical software R for distribution as an R package.

Preliminary results shown in Table 1 based on simulated data show that our algorithm describes the covariate space more aggressively (i.e., with a smaller average risk over an independent test sample [column 3]) than CART and with fewer parameters (Average Size [column 4]) than CART. This can be seen over several sample sizes (i.e., n 2 [250, 500, 1000]). All simulation results and a complete description will be available in the technical report "A Deletion/ Substitution/ Addition algorithm for partitioning the covariate space in prediction" by Annette Molinaro and Mark van der Laan.

n	Method	Average Risk	Average Size
250	ours	0.288	8.5
	CART	0.397	13.2
500	ours	0.1796	13.0
	CART	0.234	20.4
1000	ours	0.137	15
	CART	0.161	26.5

Table 1. Preliminary results based on simulated data.

Institute for Scientific Computing Research

Laboratory Directed Research and Development Project Research Summaries

LDRD Projects

The Laboratory Directed Research and Development Program (LDRD) is one of LLNL's most important vehicles for developing and extending the Laboratory's intellectual foundations and maintaining its vitality as a premier research institution. The program selects and funds an investment portfolio of high-risk, high-potential-payoff R&D projects that foster the development of new scientific and technical capabilities in support of the Laboratory's evolving national security missions. The Computation Directorate (through the ISCR) and the University Relations Program jointly oversee a portion of the LDRD program in Exploratory Research in the Institutes (ERI) that focuses on computing technologies. ERIs have the specific charter of engaging academia as a part of their research activities. This section contains the annual reports for the four ERI LDRD projects supported in FY 2003.

Parallel Graph Algorithms for Complex Networks

Edmond Chow
Center for Applied Scientific Computing-LLNL

Summary

Many scientific, technological, and social networks—such as metabolic networks, the World-Wide Web, and human interaction graphs —are being studied as "complex networks": graph topologies that appear random, but have very specific clustering and vertex-degree properties. At LLNL, a particular complex network called a "semantic graph" is being developed to organize and search information across diverse and extremely large databases. Such graph data structures will contain vast amounts of information and must be stored and queried on distributed-memory parallel computers. Our objective is to develop algorithms and research software that exploit the properties of complex networks and enable fast and efficient use of semantic graphs on parallel computers.

In FY03, we began developing a software infrastructure for semantic graphs. In particular, we have developed generators for ontologies and for semantic graphs that are consistent with these ontologies. We have developed heuristic algorithms that use ontologies to accelerate point-to-point searches in semantic graphs. We have developed a preliminary algorithm for parallelizing these searches and are beginning to understand what objectives must be optimized in partitioning semantic graphs. Further, we are applying linear algebraic techniques for clustering vertices in graphs, based on their topology. Finally, we are investigating a partitioning approach based on classifying relationships as long-range and short-range, assuming that there is a spatial embedding of the semantic graph.

In FY04, we will develop partitioning algorithms that utilize augmented ontologies, i.e., ontologies that are annotated with the frequencies of edge and vertex types in the semantic graph. The simplest approach is to partition the augmented ontology as a weighted graph and induce this partitioning on the semantic graph. However, we expect better results to be achieved if the augmented ontology is refined to have some of the major topological properties of the semantic graph. This refinement can be accomplished by clustering similar vertices or edges. In FY04, we will continue to pursue promising approaches discovered in the previous year. These include dynamic partitions based on analysts' queries. Finally, in FY04, we will begin the parallel implementation of some of our algorithms.

This work has relevance to a number of different current and future applications at LLNL. As one example, the Department of Homeland Security has many different organizations, each with its own extremely large databases of information. This technology of semantic graphs with fast and efficient search algorithms could provide a method for working with all of the data as one enormous database.

Enabling Large-Scale Data Access

Terence Critchlow
Center for Applied Scientific Computing-LLNL

Summary

This project's goal is to develop an infrastructure capable of providing scientists with access to large numbers of data sources through a single, intuitive interface. This interface will simplify scientists' interaction with data and enable them to answer more complex questions than currently possible. Our infrastructure makes extensive use of a novel metadata infrastructure to identify and describe Web-based interfaces to complex, scientific data sources. The resulting descriptions are then passed to an extended version of XWrap, a wrapper-generation program. We have developed an initial prototype of this infrastructure and have performed initial testing of the system.

This project will position LLNL as a leader in several technology areas, including data integration, bioinformatics, metadata, Web-aware agents, and wrapper generators by developing an infrastructure for accessing extensive data sources through a single interface. We plan to demonstrate the feasibility of this approach by automatically identifying sequence-similarity search interfaces across a variety of genomics Web sites.

We met all of our deliverables for this fiscal year. Specifically, we wrote the initial specification of the service-class description format, defined a new version of the interface-description format, completed the initial version of the Web spider and used it to successfully identify 17 of 25 randomly selected BLAST sites, extended XWrap Composer to accept complex interface description as input and to generate wrappers based on these descriptions, used Composer to generate wrappers for two complex BLAST interfaces, extended the DataFoundry interface to interact with BLAST wrappers, and incorporated two of the wrappers into this infrastructure.

Our milestones for FY04 are to (1) implement complex interface-identification capabilities that will support automatic identification of simple indirection pages, (2) finalize interface-description and service-class-description formats, (3) extend the Web spider to support automatic generation of interface descriptions, (4) refine Composer to accept the final interface-description format and use it to generate complex wrappers, and (5) demonstrate end-to-end automatic wrapper-interface generation capabilities for both simple and complex interfaces.

This work supports national security and other LLNL missions by benefiting ongoing programs at LLNL such as nonproliferation and stockpile stewardship, which could better utilize information from a wide variety of sources if this project is successful. Our infrastructure would simplify creating an interface that combines local data with related information publicly available over the Internet, such as scientific publications, chemistry databases, urban planning information, and census data.

Robust Real-Time Techniques for Detection and Tracking in Video

Chandrika Kamath
Center for Applied Scientific Computing-LLNL

Summary

The use of video cameras for monitoring and surveillance is becoming prevalent in diverse applications such as battlefield awareness through the identification of troop movement, national security, monitoring of traffic, detection of pedestrians, identification of anomalous behavior in a parking lot or near an ATM, etc. While a single image provides a snapshot of a scene, the different frames of a video taken over time register the dynamics, making it possible to capture motion.

The two key tasks in mining video data are the detection of moving objects and the tracking of these objects as they move over time. Once the objects have been detected and tracked, this information can be used to build models of normal behavior in the scene, enabling us to flag behavior that is abnormal. In this Laboratory Directed Research and Development (LDRD) project, we are investigating detection and tracking algorithms that are both accurate and robust (to minimize false alarms and missed positives) and real-time (to allow for corrective action). In particular, we are interested in video taken by a stationary camera under adverse conditions such as fog, or at a low resolution, or at a low frame rate. Our algorithms and software include techniques to separate the moving foreground from the background, to extract features representing the foreground objects, and to track these objects from frame to frame, followed by post-processing to smooth the tracks.

Since the start of the project in March 2003, we have identified realistic videos of traffic intersections taken under different conditions, designed a software infrastructure to handle color video, implemented several techniques for background subtraction, and evaluated their performance on our test videos. The best performance was obtained by using a mixture of Gaussians to maintain the background; however, this technique is quite expensive. A simple median of the frames also worked well and computationally was much less expensive. Further reduction in compute time was obtained by using an approximation to the median. In addition, we implemented simple schemes to calculate features such as the velocity of the moving objects.

We also conducted research on block matching techniques and found that it was effective for moderate resolution video. We could improve the performance in the presence of small camera motion by keeping a frame history. Though our main focus has been on tracking vehicles, we also supported a collaboration with the University of Colorado-Boulder on the tracking of multiple people.

In FY04, we will develop more accurate schemes to extract features such as the velocity and address the problem of tracking the objects across frames. We will also investigate how the techniques carry over to the lower resolution, lower frame rate videos.

This capability to detect and track moving objects in video supports LLNL's national security mission. For instance, this capability would be useful in counter-terrorist applications.

ViSUS: Visualization Streams for Ultimate Scalability

Valerio Pascucci
Center for Applied Scientific Computing (CASC)-LLNL

Summary

Modern scientific simulations and experimental settings produce ever-increasingly large amounts of data that traditional tools are not able to visualize in real time, especially on regular desktop computers. This inability of the scientists to interactively explore their data sets creates a frustrating slowdown in the overall process of scientific discovery.

The ViSUS project is developing a suite of progressive visualization algorithms and a data-streaming infrastructure to enable interactive exploration of large scientific datasets. The methodology optimizes the data flow in a pipeline of processing modules. Each module reads in input and writes in output a multi-resolution representation of a geometric model. This provides the flexibility to trade speed for accuracy, as needed. The dataflow is streamlined with progressive algorithms that map local geometric updates of the input into immediate updates of the output. A prototype streaming infrastructure will demonstrate the flexibility and scalability of this approach visualizing large data on a single desktop computer, on a cluster of personal computers, and on heterogeneous computing resources.

In the first two years the ViSUS project has published 19 peer-reviewed papers and produced software tools for the visualization of 2D/3D scalar fields and surface meshes. The main achievements for FY03 include:

- Completed an end-to-end prototype of the streaming infrastructure for the progressive monitoring of a Computational Fluid Dynamics (CFD) simulation (GP code).

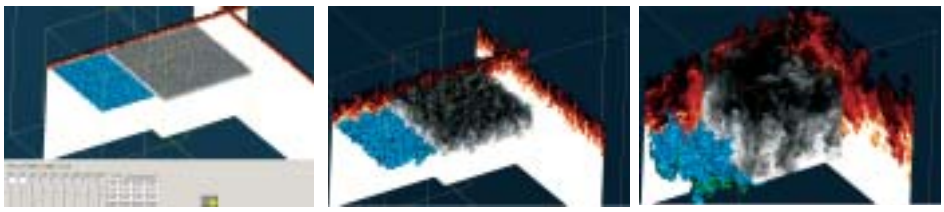


Fig. 1. Three timesteps of a Raleigh-Taylor simulation computed by the MIRANDA code (LLNL AX division) and visualized with the ViSUS Progressive Viewer. Images generated on a DELL laptop computer.

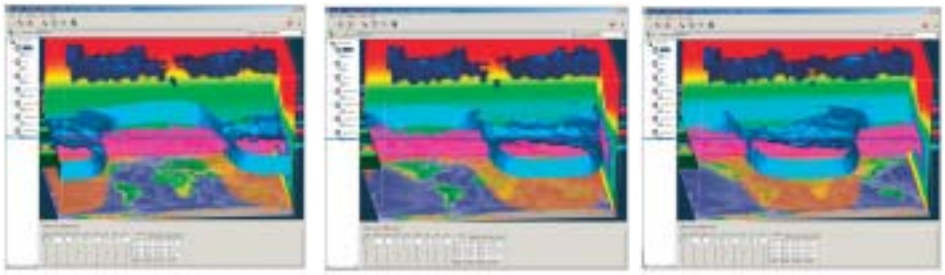


Fig. 2. Three timesteps of a climate modeling simulation computed by the IMPACT code (LLNL NCAR division) and visualized with the ViSUS Progressive Viewer. Images generated on a DELL laptop computer.

Continued

Valerio Pascucci
Center for Applied Scientific Computing (CASC)-LLNL

Summary continued

- Produced an initial prototype for the integration with the MIRANDA and IMPACT codes (see Figures 1 and 2).
- Developed an external memory technique allowing interactive traversal of multi-resolution surface meshes of arbitrary size.
- Completed the definition and prototype implementation of a wavelet model for our volumetric subdivision scheme, which will be used to extend our approach to non-rectilinear meshes.
- Completed a prototype implementation of an image cache engine that allows maintaining interactive data exploration even for slow rendering algorithms.

In FY04 we will bring the ViSUS technology to a level maturity and robustness that allows deployment to a few targeted users and demonstrates the practical advantages provided by our tools. The main milestones include:

- Develop new techniques that accelerate the isosurface construction process with occlusion culling, graphics hardware, and simple view-dependent adaptive refinements.
- Develop a new approach that allows for multi-resolution streaming of large triangulated meshes.
- Incorporate in the ViSUS Progressive Viewer the full slicing, isocontouring and volume rendering capabilities.
- Deploy the ViSUS streaming technology to the MIRANDA code (AX division) and IMPACT code (NCAR division).
- Provide a prototype viewer with new data-structures capable of handling the unstructured data format used in the HYDRA (National Ignition Facility [NIF]) simulation code;
- Provide easy installation procedures and lab-wide access to our software.

Use of our innovative, high-performance visualization techniques will allow interactive display of very large data sets on simple desktop workstations and the monitoring (or steering) of large parallel simulations. This will have specific applications to several of the DOE's and LLNL's missions, including stockpile stewardship, energy and environment, nonproliferation, biology, and basic science that use large-scale modeling and simulations. More specifically, the Laboratory will benefit from the new ViSUS technology at least at two levels. At the deployment level, the improved efficiency in the use of hardware resources will reduce the cost of visualization-hardware infrastructures. At the scientific level, the developed technology will reduce the overall time required for the design, simulation, and visualization cycle. For instance, the ability to remotely monitor large and expensive simulations will save computing resources through early termination and restart of erroneous test simulations.

Institute for Scientific Computing Research

ISCR Subcontracts Research Summaries

Subcontract Research Summaries

The ISCR oversees research subcontracts awarded to universities in the areas of Computer and Computational Science. Funding for these subcontracts can come from a variety of research budgets, almost all of which are managed by the Center for Applied Scientific Computing (CASC).

Visual Tracking of Multiple People

Michael C. Burl
University of Colorado, Boulder

Summary

Data mining has been defined as the process of extracting implicit, nontrivial, previously unknown, and potentially useful information from data in databases. Although much of the effort in the data-mining community has indeed focused on traditional databases, there is a growing interest in algorithms that can extract useful information from nontraditional data, such as video sequences. This trend is driven both by a heightened emphasis on security and by the ready availability of cheap hardware for quality video capture and analysis. It is well known that human observers have difficulty paying careful attention to massive amounts of data for any length of time. This problem is exacerbated if an operator is responsible for monitoring multiple video feeds simultaneously (e.g., a bank of display monitors). In addition, there are situations in which it is not feasible to have a human in the loop (e.g., space exploration, UAV surveillance, unattended sentinels that communicate with base only when something important happens).

In restricted domains where the environment can be controlled and the cost of errors is low or where there is a human in the loop, a few commercial video data-mining systems and advanced research prototypes that monitor people and their activities have begun to emerge. While these early systems provide an interesting degree of functionality, there is no completely satisfactory solution for the problem of autonomously monitoring activity in unstructured outdoor environments.

A significant portion of the research effort in mining video data is necessarily focused on the computer vision and image processing steps that must directly handle large volumes of raw data. The outputs from these specialized feature-extraction steps then serve as the input to higher-level data mining processes that answer questions relevant to a user. We focus on developing a prototype algorithm for extracting information from raw, surveillance-style video of an outdoor scene containing a mix of people, bicycles, and motorized vehicles.

A pilot study was conducted to develop techniques to extract useful information from image streams (video data), focusing primarily on the problem of tracking multiple people in video sequences taken from a distance. Reliable tracking is a prerequisite for supporting more sophisticated data-mining operations such as activity classification, anomaly detection, and vigilant monitoring for trigger events. A prototype algorithm, based on robust background estimation, spatial clustering, and multi-object tracking, was developed to process sequences of video frames into track sets. The extracted track sets encode the positions, velocities, and appearances of the various moving objects as a function of time. The prototype tracking algorithm was applied to a sequence of 18,000 gray-scale frames of an outdoor scene containing a mix of people, bicycles, and motorized service vehicles. The video frames were captured on the University of Colorado Boulder campus with a standard web camera

Continued

Michael C. Burl
University of Colorado, Boulder

Summary continued

situated approximately 25m above ground level. Results show that the algorithm is largely successful in tracking objects, but there are clear opportunities for further improvements.

Results also demonstrate the ability to perform data-mining queries over the extracted track sets. These preliminary studies show that useful information for surveillance/security operations or for input to public-planning and decision-making processes can be automatically obtained from raw video sequences even at the current level of tracking performance. Example utilities that that might be useful within this framework include setting up trigger events (e.g., detecting any vehicles entering a particular area, people exiting or entering a particular building), determining typical and anomalous patterns of activity, generating person-centric or object-centric views of an activity, classifying activities into named categories (e.g., walking, riding a bicycle), grouping activities into unnamed equivalence classes (clustering), and determining interactions between entities. We hope to pursue several of these directions in follow-on work.

Scientific Visualization for SAMRAI

Walter Freddy Herrera Jimenez
Oregon Graduate Institute

Summary

The goal of this project is to enhance the scientific visualization tool called VisIt to support H-Adaptive data generated by SAMRAI applications. Our enhancements are primarily concerned with supporting SAMRAI users' needs, but we intend to generalize our approaches, where possible, to support H-Adaptive data in general. H-Adaptive data is data in which the spacing between samples is varied (adapted) so that the mesh and variables may have higher fidelity (resolution) in localized areas. The kinds of work necessary in VisIt fall into roughly four categories: the GUI, the database, and I-blanking.

We worked as a group, including Peter Williams, Mark Miller, Hank Childs, and myself. My work was to develop the SAMRAI database reader and to modify AVT modules that allow VisIt to show default plots as soon as the user opens any SAMRAI database. The code was written in C++ and uses HDF5 and VTK libraries.

The current stage of the VisIt enhancement allows the user to open SAMRAI databases and use all the plots and operator tools to visualize and analyze its mesh and variables. By now, the visualization techniques consider that each level in the resolution hierarchy of the SAMRAI database is an independent mesh with its own variables. In the future, VisIt will have to "do the right thing" where multiple choices for what to render are available because of overlapping patches at multiple levels of resolution. In general, the right thing is for VisIt to display the highest resolution available in any one region of the mesh from the possible set of levels in the current selection.

Adaptivity and Related Algebraic Multigrid and Nonconforming Domain Decomposition Methods in a Massively Parallel Computing Environment

Raytcho Lazarov
&
Joseph Pasciak,
Texas A&M University

Summary

We have worked on several research projects on design, analysis, and implementation of parallel methods for various large finite-element problems. These include a least-squares finite-element method for Maxwell's equations, continuous and discontinuous finite-element approximations for second-order problems, and domain decomposition methods using matching and nonmatching grids (e.g., mortar approximations, penalty, and discontinuous Galerkin methods). These projects are related to the previous works and should be viewed as continuation and extension of the collaborative research between Texas A&M University and LLNL for the last five years. We have been working on discretization methods for partial differential equations (PDEs) that provide greater flexibility in the grid generation process, that increase the portability of various approximation methods and computer implementations, that enhance the capabilities of coarsening strategies in parallel algebraic multigrid methods, and that provide a posteriori error analysis for parallel adaptive methods. Our investigations produce competitive algorithms that can be used in various codes for complex applications in physics and engineering.

New techniques for Maxwell's equations. The goal of this project was to develop efficient algorithms for the numerical solution of problems arising from electromagnetic models. Specifically, we are interested in approximating Maxwell's equations and related eigenvalue problems. These are important in practical applications in which one needs to compute the electromagnetic field generated by prescribed current and charges or in the computation of the eigenmodes that will propagate through a given medium. The work was carried out by summer students T. Kolev and D. Copeland in collaboration with Joseph Pasciak from Texas A&M University and Panayot Vassilevski and Daniel White from Livermore Center for Advanced Scientific Computing (CASC).

One of the approaches considered is based on a weak variational formulation of div-curl systems corresponding to the electrostatic and magnetostatic problems. The finite dimensional approximation is a negative norm finite-element least-squares algorithm that uses different solution and test spaces. This algorithm allows for approximation of problems with low regularity where the solution is only in L^2 and the data resides in various dual spaces. The solution operators for the above problems are further used to obtain an approximation to the eigenvalue problem. The resulting discretization method has the advantages of avoiding potentials and the use of Nedelec spaces. A computer program was developed that applies the method to the Maxwell equations and the eigenvalue problem in the frequency domain. It is written in C++, in the framework of the AggieFEM finite element library, which supports complex geometries, local refinement, multigrid preconditioning, and OpenGL visualization. The code is based on solvers for the magnetostatic and electrostatic problems. It works on triangular, tetrahedral, and hexahedral meshes. It provides an eigenvalue solver, which allows for computations of blocks of eigenvalues and a solver for the full-time harmonic system.

Continued

Raytcho Lazarov
&
Joseph Pasciak,
Texas A&M University

Summary continued

A parallel version of this code was implemented based on MPI and CASC's Hypre preconditioning library.

A second part of the research involved the efficient computation of time-dependent Maxwell's problems. In this research, we considered a classical approach based on Nedelec curl-conforming spaces. The goal was to compare explicit time-stepping methods with implicit methods. The explicit method requires relatively little work per time step; however, the time-step size is constrained by a CFL-type condition. In contrast, the implicit method requires the solution of a system at each time step but is unconditionally stable so that larger time steps can be taken. We considered preconditioned iterative methods for solving the implicit problem with preconditioners based on algebraic multigrid. The computational results clearly showed that the implicit method with preconditioned iterative solve led to effective approximation at a significantly reduced cost on a variety of model problems.

Finally, we investigated iterative eigensolvers for the discrete equations that result from Maxwell eigenvalue problems discretized by Nedelec curl-conforming elements. The problem is to compute a modest number of the smallest nonzero eigenvalues, while avoiding the large null space of the curl–curl operator. We investigated a number of techniques for doing this. The most effective was a block iterative approach with preconditioning for the operator curl–curl plus the identity, while approximately enforcing the zero divergence condition. Initially, the zero divergence condition was approximately imposed by a preconditioned conjugate gradient iteration. Subsequently, this condition was imposed using the conjugate gradient iteration without preconditioning.

Domain Decomposition Methods for Mortar and Nonmortar Approximations. The interior penalty method aims at eliminating the need for additional (Lagrange multiplier or mortar) spaces and imposes (only approximately) the required continuity across the interfaces by an appropriate penalty term. In our approach, the jumps in the values of the functions along these interfaces are penalized in the variational formulation. For smooth solutions, we lose the optimal accuracy due to lower approximation at the interface, but we produce symmetric and positive definite discrete problems that have optimal condition number. Other features of our method are: natural parallelization of the iterative procedure and parallel mesh generation based on adaptive procedure that involves local-error estimators and indicators.

We also continued our research into using the mortar method for patching together nonconforming meshes that result when algebraic multigrid (AMG) methods are applied independently on subdomains distributed in a parallel computing environment. We have developed a code that implements the above technique using the AMGe spectral

Continued

Raytcho Lazarov
&
Joseph Pasciak,
Texas A&M University

Summary continued

agglomeration technique of Panayot Vassilevski. This code works for three-dimensional second-order model problems and develops the necessary parallel data structures for implementation of the global AMG preconditioner. Currently, the overhead associated with setting up the subprocessor agglomeration is the limiting factor in a moderately parallel application. To circumvent this problem, we propose to extend the approach by using other agglomeration techniques. We also propose to test the approach on machines with many more processors. A parallel code implementing the mortar method with algebraically constructed multiplier spaces was developed. The target application of this code is in the construction of parallel (including multigrid) preconditioners using element based (AMGe) coarsening in each subdomain. The work was carried out in close collaboration with our summer student T. Kolev from Texas A&M University and P. Vassilevski from CASC.

Discontinuous Finite-Element Approximation of Elliptic Problems. Jointly with P. Vassilevski we have proposed a framework for derivation of discontinuous finite-element discretizations of elliptic problems on matching and nonmatching grids and possibly many subdomains. The derivation of the schemes is based on penalty stabilization of the mixed formulation that allows elimination of the Lagrange multiplier. In particular, many known discontinuous finite-element Galerkin methods can be reproduced by the proposed framework. However, this could be considered as a way to develop approximations and domain decomposition methods for nonmatching grids. Stability, error analysis, and preconditioning methods for solving the discrete system of equations are developed. The results are in further development. We implemented the following three discontinuous Galerkin methods: the method of Baumann and Oden, the nonsymmetric interior penalty Galerkin (NIPG) method, and the interior penalty (IP) method.

The code was parallelized using MPI. All matrices are stored in Hypre ParCSR format, and the Hypre implementation of the iterative methods is used. The code has the following structure: (1) the root process reads an arbitrary tetrahedral mesh from a file and then refines it uniformly a given number of times; (2) the root process uses METIS to partition the new mesh into the number of subdomains equal to the number of processors used; (3) the root process sends to each processor its corresponding submesh (all elements in its subdomain), which will be the coarsest-level mesh for the multigrid; (4) the mesh is refined uniformly a given number of times in parallel (when multigrid preconditioner is used, at each level we assemble in parallel the matrix of the arising linear system, the interpolation operator, and the smoother used); and (5) on the finest level, the right-hand side is assembled and then the GMRES (Baumann–Oden and NIPG methods) or the PCG (for IP method) algorithm is used to solve the linear system iteratively.

Continued

Raytcho Lazarov
&
Joseph Pasciak,
Texas A&M University

Summary continued

We use a pure geometric multigrid method, with operators obtained by assembling at each level, thus generating (in general) nonnested bilinear forms. The natural embeddings of the discontinuous spaces were used as interpolation operators, thus giving rise to purely local interpolation–restriction. In the multigrid algorithm, the following two smoothers were considered:

• diagonal scaling—the preconditioner has elements on the main diagonal equal to the sum of the absolute weight of the rows. With this smoother, a variable V-cycle was used with three pre- and postsmoothing iterations on the finest level, and the number of smoothing iterations is multiplied by 4 from a level to the next coarser level;

• the ParaSails preconditioner from the Hypre library with the default settings. For the IP method, the SPD version of ParaSails was used. Here, V-cycle with one pre- and postsmoothing iteration was used.

Interaction with CASC members, visitors, and summer students. Raytcho Lazarov and Joseph Pasciak visited CASC for the month of August 2003. They were also in close contact with J. Jones, D. White, R. Falgout, C. Tong, and P. Vassilevski and interacted with other long- and short-term visitors R. Bank, T. Manteuffel, S. McCormick, C. Bacuta, and L. Zikatanov. Our scientific collaboration for this year resulted in joint papers that are submitted or already published. Three Ph.D. students from the program of computational mathematics at Texas A&M University, T. Kolev, V. Dobrev, and D. Copeland, spent three months each at CASC on a professional summer internship. All three Texas A&M Ph.D. students participated directly in the work of the contract. We plan to develop, implement, and test efficient parallelizable, scalable algorithms on various problems for treating both the geometrically refining spaces and algebraically coarsening grids. The future research should offer alternatives to the existing efforts in CASC for parallel AMGe method.

Enabling Large-Scale Data Access

Ling Liu
Georgia Institute of Technology

Summary

An ultimate goal of the Enabling Large-Scale Data Access project is to develop methodology and mechanisms for building a fully automated, end-to-end wrapper code generator. This can be achieved through the design, development, and integration of service-class descriptions with the XWRAP systems. The main idea is to provide mechanisms to enable XWRAP systems to take a generic description of a class of search interfaces (so-called service class) and the URL of a particular interface that is a member of the given class to produce a functional wrapper; this wrapper will take a class-specific query and produce an XML view of the query results obtained through this particular interface. In the second year of the Laboratory Directed Research and Development (LDRD) project, we focus on three main efforts: (1) developing concrete syntax for specifying the service-class descriptions, (2) designing and implementing a spider for surfing the Web and finding those Web sites that match with a given service class, and (3) integrating the service-class-based spider with the XWRAP system to automate the code-generation process of XWRAP.

We first give an overview of the work performed under the LDRD subcontract by the Distributed Data Intensive Laboratory (DiSL) at the College of Computing in Georgia Tech. We present the overview in terms of the deliverables defined in the statement of work.

Deliverable 1: *Java 1.3 code for a Web spider capable of taking a service-class description and a starting URL, crawling the Web, identifying interfaces that support a simple BLAST query, and generating a description of how to interact with that interface in the format required by the XWRAP Composer program described* in Deliverable *(3).*

Deliverable 2: *An XML service-class description for a BLAST query.*

Deliverable 3: *A JAR file containing a demo version of the wrapper program generated by XWRAP Composer, which is capable of taking the service-class description generated by the spider referenced in Deliverable (1) and producing a java wrapper for that source.*

Deliverable 4: *A short whitepaper describing the interface description format output by the spider in Deliverable (1) and used by XWRAP Composer (3).*

Deliverable 5: *A final report describing the work completed during this subcontract, any problems that occurred, possible extensions to the system, and how the software can be installed and used.*

We have completed all the tasks described in the above deliverables and report briefly selectively here.

Service-Class Description and Source Discovery. One of the main results of our second-year effort is the development of the service-class descriptions for a selection of Bioinformatics Web sources and the service-class-driven spider that is capable of

Continued

Ling Liu
Georgia Institute of Technology

Summary continued

crawling the deep Web and discovering the sources that match the given service-class description.

Concretely, our approach to discovery and classification of Web sources groups them into *service classes* that share common functionality but not necessarily a common interface. Service classes are specified by a *service-class description*, which uses an XML format to define the relevant aspects of a category of Web sources from an application's perspective. The service-class-description format supports the source-discovery problem by providing a general description of the type of source that is considered interesting. It defines the data types that comprise the service arguments as well as any intermediate types that may appear in a source. It establishes a general description of the interface used by source-class members and outlines intervening control points. Finally, it lists examples that are employed during source evaluation. Each of these components is described in detail in the remainder of this section.

The first component of a service-class description specifies the data types that are used by members of the service class. Types in this context are analogous to those in programming languages. They are used to describe the input and output parameters of a service-class and any data elements that may be required during the course of interacting with a source. The control flow graph consists of a set of states connected by edges. Each state has an associated type; data from a Web source is compared against the type associated with the control-flow states to determine the flow of execution of a source from one state to another.

The service-class description examples provide the mechanism by which a source can be analyzed and tested for service-class membership. Examples are tied to one or more paths through the control flow and specify data for the input types at a particular control-flow state; there may be many examples specified in a service-class description. Each example consists of a set of arguments that are paired with form parameters on the Web source. An argument has a name, a data type defined in the types section, a set of hints, and a data value. Hints leverage the observation that because Web pages are constructed by humans, form parameter names and values tend to reflect their purpose. The hints provide clues that allow the source analyzer to tie an argument to form parameters with a better chance of picking the correct parameter.

Source Analyzer. We have constructed the code to accept a service-class description as input and to use that description to determine service-class membership for a given Web source. Source analysis encompasses the following steps:

- input parsing an object materialization,
- test-query generation, and
- query execution and response analysis

Continued

Ling Liu
Georgia Institute of Technology

Summary continued

During input parsing, the source analyzer reads in the service-class description and constructs a set of objects that represent that description and provide useful methods for operating on data obtained from a Web source. Our work here was to construct the objects that represent the service-class in memory along with the desired functionality for those objects. The mechanics of parsing a service-class description into in-memory objects are handled by the Apache Digester library, which takes an XML file and a set of rules and generates arbitrary Java objects according to the specifications in the rules. The source analyzer also retrieves the start page of the supplied Web source during input parsing. Network communications and HTML handling are provided by the HTTP Unit programmatic-user agent library.

Once the input data has been processed, the source analyzer uses the examples in the service-class description and the forms from the Web source's start page to generate a set of queries that will be used to probe the source. The analyzer's goal is to discover the function of the various inputs on the form and thereby determine if the source matches the service-class description. Queries are generated by computing all possible permutations of the example arguments with the form parameters. From this set, queries are chosen to be executed in an order determined by a scoring function that gives higher priority to queries where the form parameter chosen for a particular argument matches the hints specified by that argument. In this way, the source analyzer tries to execute queries that it deems more likely to succeed first.

After generating the list of queries to execute, the source analyzer begins probing the site and testing the responses. Response analysis is handled by the type library. A Web source's response is tested by the type library using the type defined for the current control state in the service-class description. If the response matches the specified type, the analyzer progresses to the next control state; if the current control state is an end state, the Web source is declared to be a match for the service-class description and processing concludes.

XWRAP Composer. XWRAP systems aim at semi-automatically generating Java code (Wrapper) for extracting useful content from Web pages. XWRAP Elite is designed to generate wrapper programs for extracting information (primarily query-answers) from the deep Web data sources. XWRAP Composer is designed on top of XWRAP Elite, aiming at providing information extraction across multiple linked Web pages. A typical example is the Matt's scenario [3].

The XWRAP Composer requires three types of input descriptions to generate the Java code of the given Web source from the service-class specification.

- The first input description is called Interface Description, which is used to specify the URL of the Web source to be wrapped and the keyword or entry query used to obtain the pages from the search interface of the given Web source.

Continued

Ling Liu
Georgia Institute of Technology

Summary continued

- The second input description is called Output Service-class Description. It describes the XML tagging information and how extracted data content can be tagged as specified.
- The third input parameter is the Composer Script Language, which consists of two main components: the QA control logic specification and the data extraction logic required. The QA control logic specifies the possible query-answer control flows of the search interface of a given Web site. An example of such control logic is as follows. The NCBI BLAST interface will respond to a BLAST query with at least three different control flows:
  - The first one is the waiting page, which tells how long one needs to wait for the results to be returned.
  - The second one is the full result page, which is returned by following the indirection link on the intermediate waiting page.
  - The third one is the error page, which is returned when following the indirection link on the waiting page.

Thus the Composer wrapper needs to generate code pieces to handle all three cases.

Generating a script definition for XWRAP Composer from the source analyzer involved the creation of code to further analyze a discovered Web source and to tie the various pieces of information obtained from the service-class description and query probing into a form that Composer could understand. Additional site analysis was required in order to determine fully the capabilities of a recognized Web source. For example, in BLAST, it is possible for a single source to provide a gateway into many databases, each of which should be available in the system and thus require wrappers. The additional site analysis is similar to the site analysis used when initially testing the source, but involves more exploration of the various options presented by the Web source's forms.

Once the analyzer has obtained all of the necessary information, script generation can begin. Each script is divided into several tasks. The source analyzer begins by constructing a task that will read in the service-class description to provide later tasks with type definition and other data they will need to complete their work. The second task created by the source analyzer takes input from a user in the system and constructs the messages needed to submit the query to the Web source. Next, the script specifies a task that will annotate the Web source's results with XML tags. The final task takes the annotated results and constructs data objects that will eventually be passed to the user.

The generated XWrap Composer script follows the interface specification laid out in the attached specification document and must conform to the schema definition. Two versions of the schema definition are provided: one in RelaxNG compact syntax for easier readability and a standard XSD description generated from the compact syntax. For tools to make the translation, see http://thaiopensource.com/relaxng/.

Continued

Ling Liu
Georgia Institute of Technology

Summary continued

Results Obtained from Integration Efforts. One of the main objectives of Deliverable (3) is to produce a design document describing how the service-class description of a remote data source generated by the spider referenced in Deliverable (1) is fed into the XWRAP Composer to generate wrappers in Java for that source without any human involvement. A demo is included to show the feasibility of such design.

There are two challenges for providing a fully automated, end-to-end wrapper generator system. First, we need mechanisms that can provide all the information required by XWRAP code-generator system through human interaction with XWRAP. To our knowledge, all wrapper-generator systems to date are semi-automated, and they require wrapper developers to enter information to the code-generator systems at different stage of the code-generation process.

In XWRAP. there are three manual steps that require human input in its code-generation process.

- First, it requires a manual input of the URL of the Web site to be wrapped.
- Second, it requires manually writing the Composer script to capture the control logic of the query-answering patterns of the Web site to be wrapped.
- Third, it requires manually inputting the data-extraction complications such as element alignments, or multi-page extraction through links.

The second year of the LDRD project focuses on mechanisms that can provide full automation by utilizing the manually developed service-class description to capture all the human-input information required by the XWRAP code generator. This effort is carried out in three steps. First, we manually create the service-class description tailored to the type of Web sites to be wrapped such as BLAST sites. Then, we find the Web sites matching a given service-class description. Finally, we integrate the service-class description with the XWRAP Composer, such that the service-class description for XWRAP Composer will provide interface, outerface description as well as the QA control logic and the data-extraction logic required by the XWRAP Composer code generator.

Concretely, the service-class-description-based spider is used to discover a set of Web sites that share a similar search interface such as BLAST. Thus, instead of entering a specific URL to the XWRAP Composer, one can simply request for a wrapper to be generated capable of providing some specific services such as BLAST. The service-class description is used to find the concrete Web site offering the type of services — for example, the BLAST sites. For each of such Web sites, its URL can be passed to the XWRAP for generating a wrapper for that site.

Continued

Ling Liu
Georgia Institute of Technology

Summary continued

It is observed that many Web sites share similar search interfaces and QA control logics [2]. Therefore, by including the manual specification of the QA control logic required by XWRAP to wrap a Web site in its matching service-class description, it makes it possible to automate the generation of the QA control-logic part of the XWRAP Composer script.

For the data-extraction logic, Web sites offering the same types of services tend to differ in how they organize and lay out the presentation of the content. Therefore, by moving some type of data-extraction logic into the service-class description, a certain number of Web sites that provide the same type of services will not be discovered due to the mismatch in terms of their data-extraction logic.

Our experience with the integration efforts yields two observations:

(a) the types of BLAST sites where service classes work well may cause the object-extraction algorithms used in Omini to fail; thus, XWRAP will not be able to generate wrappers, and
(b) Web sites that XWRAP/Omini can work well may not be the best Web sites for the service-class spider (such sites will need more complex interactions between wrapper developer and the XWRAP system).

One possible solution is to add what XWRAP needed to generate the Entrez wrapper (the alignment rules) to our Entrez service-class specification. This will allow us to integrate the two pieces of code, which takes the input to the service-class spider and returns an Entrez wrapper. Another solution is to find those Web sites that a service-class generator can work more or less correctly, and Omini will also be able to find the correct objects.

There are three papers directly published in this project.

- The first paper [1] reports our work on source discovery and classification. The claimed contribution of the paper is the use of an abstract description format for the classification of arbitrary Web sources. The paper describes the service-class description, the source analyzer, and the validation results obtained by testing the code against a variety of Web sources. Using BLAST as a case study, we showed that the source analyzer is able to correctly identify 66% of the supplied BLAST sources with no false positive identification of non-BLAST sites. The paper was accepted to the Atlanta Bioinformatics conference and is scheduled for publication in November 2003.
- The second paper reports our new data extraction algorithms [2] which appeared in the IEEE International Conference on Data Engineering with *a* 14% acceptance rate. This paper presents QA algorithm for focused extraction.
- The third paper [3] summarizes the issues involved in data integration over more than 500 Web sources.

Continued

Ling Liu
Georgia Institute of Technology

Summary continued

Publications

[1] Daniel Rocco and Terence Critchlow. *Automatic Discovery and Classification of Bioinformatics Web Sources.* In Proc. Atlanta Bioinformatics Conference. 2003 (to appear)

[2] James B. Caverless, Ling Liu, and David Buttler. *Probe, Cluster, Discover: Focused Extraction of QA-Pagelet in Deep Web Data Sources*, To appear in IEEE International Conference on Data Engineering. 2004.

[3] David Buttler, Matthew Coleman, Terence Critchlow, Renato Fileto, Wei Han, Calton Pu, Daniel Rocco, Li Xiong. *Querying Multiple Bioinformatics Information Sources: Can Semantic Web Research Help?* SIGMOD Record, Vol. 31, No. 4. 2002 (December)

Optimizing a Parallel Code for the Simulation of Neuronal Networks

Peter Pacheco
University of San Francisco

ISCR Institute for Scientific Computing Research

Summary

Object-oriented NeuroSys is a collection of parallel C/MPJ programs for simulating very large networks of biologically accurate neurons. It includes two principal programs: ooNeuroSys, a parallel program for solving the large systems of ordinary differential equations (ODEs) arising from modeling the interconnected neurons, and Neurondiz, a parallel program for visualizing the results of ooNeuroSys. Both programs are designed to be run on clusters and use the MPJ library to obtain parallelism.

NeuroSys was originally developed in the late 1990s by the University of San Francisco's Applied Mathematics Research Laboratory. It obtained parallel efficiencies of better than 90% on networks of 250,000 Hodgkin-Huxley type neurons and a fast-ethernet-connected cluster of 32 Intel processors running Linux. However, its design is essentially structured, and as it grew, it became very difficult to incorporate new features and to improve existing ones. So in the summer of 2001, we began a complete rebuilding of NeuroSys. One of the central features of the new system is an object-oriented design that makes maintenance relatively easy and actually improves both performance and scalability. Because of the relative ease with which C codes can be optimized, we chose to write ooNeuroSys in C. Neurondiz is divided into a computational engine or "backend" and a "frontend" for managing the display. The backend is also written in C, and there are two versions of the frontend: one written in Java for maximum portability; the other written in C++ using the Qt and OpenGL libraries for maximum performance.

During October 2002–May 2003, we worked on a number of approaches to improve the performance and usability of ooNeuroSys. Perhaps the most significant improvement in the overall performance for ooNeuroSys was obtained by replacing fairly simple, nonscalable parallel ODE solvers with parallel CVODE, a code developed by researchers in the Center for Applied Scientific Computing (CASC). Another significant enhancement came in the development of adaptive methods for structuring the interprocess communication. A third avenue to performance improvement was obtained by selectively exposing incomplete data structures and replacing function calls —especially accessor functions — by macros. A fourth avenue for performance enhancement has come from the improvement of the I/O schemes used in the program.

From a computer scientist's point of view, one of the more interesting aspects of ooNeuroSys is that the interconnection networks in the systems it models range from being very sparse to very dense. Thus, it is impossible to know the best algorithm for interprocess communication until runtime. In order to address this, we first developed several different algorithms for interprocess communication. However, we found that only two sufficed for the best performance: one for densely interconnected networks and one for sparsely interconnected networks. At runtime we use a heuristic scheme to

Continued

Peter Pacheco
University of San Francisco

Summary continued

evaluate the density of the interconnection network. If the interconnect lies near the "crossover point" from sparse to dense, we run a small benchmark of etch scheme to empirically determine which interconfleet is superior. Since the interconnects change infrequently if at all, this approach adds very little to the overall cost of running the heuristic, and the benchmark is more than made up for by the overall reduction in the cost of communication.

Although much work has been done in developing sophisticated software for interfacing to *parallel* filesystems, little has been done to improve the performance of parallel software for programs that interface to conventional Unix I/O. This is an important problem since many, if not most, Linux clusters use multiprocessor nodes with dedicated local disks, and most of these systems simply use the ext2 or ext3 Linux filesystem. Thus, one of the main foci of our work has been optimizing the I/O performance of multiprocessor, especially dual-processor, Linux systems running parallel programs. Our research has shown that in order to optimize I/O-intensive programs running on such systems, it is necessary to use threads. However, for programs that make less intensive use of the I/O subsystem, we have found that a pure MPI approach combined with the use of the mmap() and msync() system calls provides excellent overall performance—performance as good as, or even somewhat better than, a mixed MPI/Pthreads approach.

In order to improve the usability of ooNeuroSys, we added a Python interface. This addresses one of the most challenging problems we faced in the original version: the difficulty of converting a system of ODEs modeling a neuron into a collection of C functions with the proper interface to the rest of the program. The Python interface is much simpler (it requires only a nodding acquaintance with programming) and because Python is interpreted, users get almost instant feedback on both the correctness and the usefulness of their model code.

We developed the frontend-backend design of Neurondiz because we assumed that our users would mainly have access to computing systems with two basic configurations. The first configuration assumes the user has access to a parallel computer with essentially no graphics capability. For example, a user who accesses a remote parallel system at a supercomputer center via the internet would fall into this category. The second configuration assumes the user possesses a parallel system in which one of the compute nodes is directly connected to the display. For example, a user with a small cluster might use this setup. For users with the first setup, the backend, the compute-engine, communicates with the frontend, the display management software using sockets. For users with the second setup, the frontend and backend are, effectively, part of the same program and they communicate via function calls. Currently our Java frontend supports sockets for communicating with the backend, and our C++ frontend

Continued

Peter Pacheco
University of San Francisco

Summary continued

supports direct function calls. We plan to ultimately implement both designs with both languages, using JNI for the Java frontend.

During the period of the LLNL visit, our main focus with Neurondiz has been on adding features to the user interface. The original version simply opened a window and stepped through the results generated by the ODE solver. It displayed a rectangular grid of disks — one disk for each neuron — and through color changes in the disks showed the changes in neuron membrane voltages. We have added the capability to "blow-up" a subset of the display for more detailed information. The new version also allows a user to select an individual neuron and display a membrane voltage vs. time plot for the neuron.

The neuroscientists we worked with asked us to provide information on the synaptic interconnections between the neurons. After some discussion we added two functionalities. The first allows the user to select a subset of neurons and display the synaptic interconnections among the selected neurons. The second allows the user to select a source and a destination neuron and a bound on the number of synaptic connections. The software then finds the subnetwork consisting of all paths of synaptic connections from the source to the destination that have length less than or equal to the bound. Both functionalities make use of the graph visualization software package, Graphviz.

For the near future, we plan to fully integrate all the Neurondiz functionalities into both the Java and C++ frontends. We also plan to further optimize the I/O and communication of ooNeuroSys. For the longer term, we plan to integrate the use of graph partitioning software into ooNeuroSys and Neurondiz.

The work done at LLNL has provided tremendous improvements in both the performance and usability of Object-oriented NeuroSys. We expect that computational neuroscientists will benefit greatly from these improvements. We also expect that the wider community of users of parallel scientific computing will benefit from much of this software and many of its design features.

Complex Networks

Alex Pothen
Old Dominion University

Summary

Graphs in which the average distance between a pair of vertices is small, i.e., O(log |*V*|), where |*V*| is the number of vertices, are called *small-world* graphs or networks. Such networks are of interest currently since they arise in social, biological, and web-community networks. Small-world graphs can be classified into several classes: Erdos-Renyi random graphs, linearized chord diagram (LCD) random graphs, and the Strogatz-Watts-Kleinberg (SWK) class of graphs.

In Erdos-Renyi random graphs, each edge is present with a probability *p(n)*, where *n* is the number of vertices. These graphs have been well studied in the literature, e.g., Bollobas' book on Random Graphs. It is well known that these graphs cannot be partitioned well in the sense that they cannot be partitioned into subgraphs with approximately equal numbers of vertices such that few edges join the different subgraphs.

LCD graphs have been studied by Barabasi and others, and we follow the construction of Bollobas. We build the graph through a random process, beginning with, say, a clique on three vertices. At each stage, we add a new vertex, joining it to two vertices in the current graph, with the probability of the endpoint proportional to its current degree. Thus, it is more probable that a new edge will join the new vertex to a previous vertex of high degree, than to a vertex of low degree. The LCD model leads to graphs whose vertex degrees satisfy a power-law, i.e., $N(d) = Ad^{-B}$, where *N(d)* is the number of vertices of degree *d*, and *A* and *B* are constants. At the beginning of this project, we had hoped that it might be possible to partition such graphs by removing the few vertices of high degree from the graph initially, partitioning the residual graph, and then adding the removed vertices to appropriate subgraphs. We implemented such an approach, but it did not give good partitions: a constant fraction of the edges in the graph were cut by the partition. In hindsight, this result was to be expected. Recently, Bollobas and Riordan have shown that the LCD model generates graphs with almost the same properties as the Erdos-Renyi random graph model. Hence it should be possible to prove that the LCD model leads to graphs that do not have good partitions similar to the ER random graphs.

The SWK graphs begin with a rectangular grid graph (in two dimensions) or a cubic grid graph (in three dimensions), with each vertex joined by edges to its neighbors in the north, south, east, and west in the two-dimensional grid (and similarly for the 3D grid). We now add a set of random edges, by choosing the endpoints at random. If we choose the pairs of endpoints without replacement, then at most one random edge can be incident on each vertex in the graph. Assume that we wish to partition the graph into *P* subgraphs. Then the probability that the random edge joins two vertices in the same subgraph is *1/P*, and that it will be cut by a partition is *(P - 1)/P* . Hence we expect that most of the random edges will be cut. We classify the random edges into two classes: *local* edges that join

Continued

Alex Pothen
Old Dominion University

Summary continued

vertices that are close to each other in the grid, and *remote* edges that join vertices that are distant in the grid.

To make this classification precise, we need to define the concept of distance in SWK graphs. The usual definition of the distance is the number of edges in a shortest path between two vertices in a graph. We define a *grid distance* of two vertices in SWK graphs as the number of edges in a shortest path that includes only edges from the grid and none from the random edges. Thus, the two endpoints of a random edge could have a large grid distance between them if the random edge is a remote edge.

We now describe an algorithm that has proved to be successful for computing a good partition for SWK graphs. We do not use information about which edges are original grid edges, and which have been randomly added, in the algorithm. The idea is to seek to identify the remote edges, remove them from the graph, and then to partition the residual graph. We identify a remote edge by temporarily removing the edge, computing the distance between the two endpoints of the edge, and classifying it as remote if it is larger than a threshold value. There is no guarantee that the distances obtained in this manner correspond to the grid distance (as we have defined it earlier), because random edges could be included in the paths we consider. But since random edges are chosen without replacement of endpoints, at most one random edge can be incident on each vertex in the graph, and with high probability we identify most of the remote edges. It is also clear that local edges will be correctly classified in this approach. (Even if we add random edges with replacement of endpoints, the probability that two or more random edges are incident on the same vertex is low, and this approach works in practice.)

The algorithm for partitioning an SWK graph then is to consider each edge in turn, compute the distance between its end points using paths that do not include that edge, and then temporarily remove the edge if it is classified as remote. After removing edges that have been classified as remote, we then partition the residual graph, using METIS, one of the well-known graph partitioning packages. Computing the distance between the endpoints of an edge can be done in $O(|E|)$ time, where $|E|$ is the number of edges, using a breadth-first-search (bfs). The algorithm can be implemented in $O(|E|^2)$ time, since $|E|$ edges need to be classified as remote or local. In practice, of course, a bfs can be terminated as soon as we reach the second end point of an edge from the first end point, and hence the algorithm runs quite fast relative to the partitioner.

We have generated a test graph consisting of a 100x100 grid, with a five-point stencil (each vertex connected to its four neighbors on the grid), and then added, in 99 steps, 100 random edges in each step. In the following two figures, we show the result of partitioning the final graph with 9900 random edges added, first using METIS directly, and then using the new algorithm that identifies and removes the remote edges before partitioning it

Continued

Alex Pothen
Old Dominion University

Summary continued

with METIS. The figures show quite clearly that the random edges confuse the multilevel partitioner METIS, which then computes a partition into two sets that loses the notion of locality in the grid graph. On the other hand, once most of the remote random edges have been identified and removed, then the partitioner partitions the grid graph into two subgraphs with good locality properties.

We believe that these results are of interest to community networks, where a few individuals serve to reduce the distance between two sub-communities that otherwise have little overlap. Results in the random graph literature suggest that a few such edges reduce the diameter of random graphs to O(log |V |). Hence the partitioning algorithm described here should be useful to partition large-scale community graphs for distributed computations.

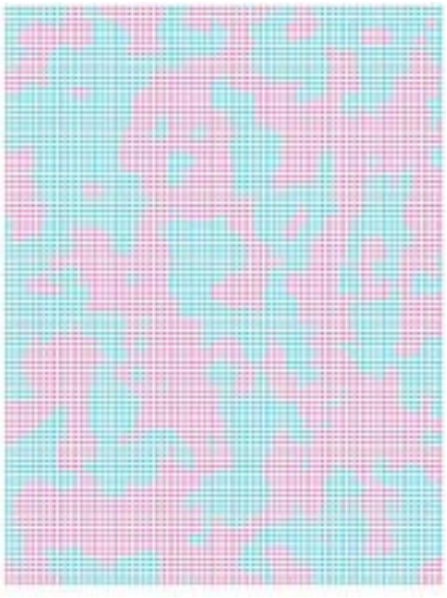


Fig. 1. The 100x100 (five-point) grid with 9900 edges added at random partitioned with METIS.

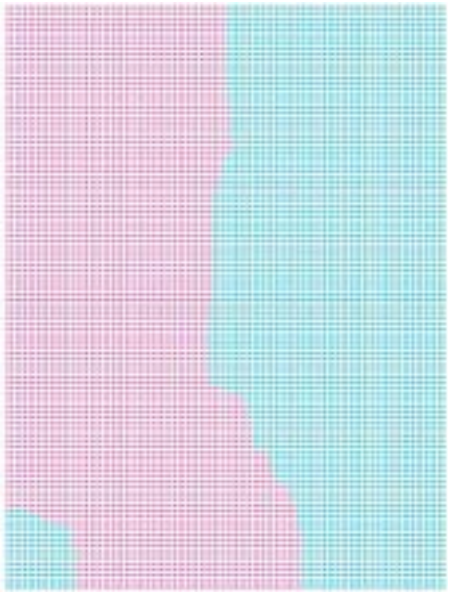


Fig. 2: The 100 x 100 (five-point) grid with 9900 edges added at random partitioned, with the new algorithm. The algorithm first identifies remote edges, removes them from the graph, and then uses METIS to partition the residual graph.

FOSPACK

John W. Ruge
Front Range Associates

Summary

The goal of this subcontract is continued development of the FOSPACK code, which is a package developed by John Ruge for the automatic discretization and solution of First-Order System Least-Squares (FOSLS) formulations of partial differential equations. Both 2D and 3D versions have been written, although the 3D version is more rudimentary. FOSPACK takes a user-specified mesh (which in 2D can be an unstructured combination of triangular and quadrilateral elements) and specification of the first-order system, and produces the discretizations needed for solution. Generally, all specifications are contained in data files, so no recompilation is necessary when changing domains, mesh sizes, problems, etc. Much of the work in FOSPACK has gone into an interpreter that allows for simple, intuitive specification of the equations. The interpreter reads the equations, processes them, and stores them as instruction lists needed to apply the operators involved to finite element basis functions, allowing assembly of the discrete system. Quite complex equations may be specified, including variable coefficients, user defined functions, and vector notation. The first-order systems may be nonlinear, with linearizations either performed automatically or specified in a convenient way by the user. The program also includes global/local refinement capability. Solution of the linear systems is performed with an algebraic multigrid solver, which is very well suited for such problems.

The specific tasks to be performed were as follows:

1. Continue work on both 2D and 3D FOSPACK codes to improve modularity of the program and generality of meshes allowed.

2. Continue work for allowing higher-order elements, with the ultimate goal of a full h-p refinement algorithm.

3. Implement and test methods for dealing with time dependence.

4. Implement and test methods for incorporating added algebraic constraints (e.g., slide surfaces).

5. Work toward parallelization of the FOSPACK code.

6. Work on development and incorporation into FOSPACK of alternative solution methods, particularly self-correcting Algebraic Multigrid [AMG] and smoothed aggregation.

7. Continue exploration of the use of FOSLS in problems of interest to LLNL personnel.

Continued

John W. Ruge
Front Range Associates

Summary continued

FOSPACK was originally written as a stand-alone code, meant for easy testing and solution of FOSLS formulations. While it was somewhat modular (with clear interfaces between the four main tasks: equation interpretation, domain/mesh definition, assembly of the linear systems, and solution of linear systems), it was never meant to be either fully modular or a library of routines with clearly defined interfaces. However, several factors make improving modularity an important goal. The first is simply that the amount of planned work is too great to be accomplished practically by one person. Efficient division of labor requires a clearer separation of various code parts, along with clearly specified interfaces. The second factor is that, as the code develops, the application to a wider range of problems, and more realistic problems, becomes necessary. The capabilities of some parts of the code, particularly the mesh generation, are not sufficient for such generality. Use of existing packages requires a clearly defined interface suitable for inclusion of such packages. The use of alternate solvers, discussed later, also requires a suitable interface.

Issues concerning these two factors are being addressed separately as they arise. Much of the emphasis on division of labor concerns the introduction of higher-order elements, discussed in more detail below. While use of third-party software for such elements was considered, there seemed to be no clear way to separate the definition of finite element space from matrix assembly. Instead, a number of tasks were defined that were required in matrix assembly, functional computation, etc., with interfaces specified. This allowed the matrix assembly and functional computation routines to be rewritten with no dependence on the function space chosen and knowledge of the spaces contained only in this new set of routines.

For mesh generation, a number of packages were examined, and a number of issues were considered, including generation of the initial mesh, refinement, and derefinement. It was decided that the best approach was to use (and enhance) the refinement and derefinement capabilities already in FOSPACK, while allowing for an initial mesh produced by external mesh generation packages. Such a capability is already partly included, and a fairly clear interface exists. Some restructuring of the code was performed, allowing for reading a “raw” mesh (list of nodes and elements along with an element-node correspondence table). Additional information (such as boundaries) can be specified in separate files, and conversion into the form required by FOSPACK is also performed separately.

Much of the work during this performance period was directed toward the introduction of higher-order elements. Currently, this is included in the 2D version only, although the routines developed were structured in such a way as to allow fairly straightforward extension to the 3D case. Higher-order functions are restricted to quadrilateral (not triangular) elements and can essentially be any desired order (although there are

Continued

John W. Ruge
Front Range Associates

Summary continued

practical limitations, such as storage and the slower convergence behavior of the AMG solver as the order is increased). The coarsest mesh problem is defined using bilinear elements, and higher-order functions are introduced as a refinement level (in a process called "p-refinement"). Initially, only global p-refinement was allowed, and only on a mesh without local h-refinement (where h-refinement is the usual element subdivision). Work is just concluding allowing for arbitrary p- and h-refinement (where the determination of p- or h-refinement can be made separately for each element). The next stage is to determine which refinement strategies are beneficial or practical.

Time stepping has been added to both the 2D and 3D versions of FOSPACK. Currently, this is implemented by explicitly writing the time-dependent terms in the first-order system in difference form (e.g., $(u-u_{old})/dt$), where u_{old} is the saved solution from a previous time step or the user-supplied initial condition. For this, a new function called *prev* was introduced, so that the time-dependent term could simply be written as *(u-prev(u))/dt*. FOSPACK then knows to allocate space and store the values of *u* at the end of each time step. FOSLS is then applied to this system, resulting in backward Euler time stepping. The time step *dt* used can be specified in several ways and can either be fixed throughout the solution process or can be adjusted by the user based on the evolving solution. (The latter requires some user coding.)

The addition of algebraic constraints has not yet been fully implemented, although little additional coding is needed. A nice method allowing the user to specify such constraints is needed.

Work towards parallelization of the FOSPACK code is continuing. As a first step, the existing AMG solver has been replaced with a call to BoomerAMG through the HYPRE interface. Comparisons with the serial AMG solver are currently in progress. In addition, the matrix assembly and functional computation are being studied for parallelization to determine whether existing code can be easily modified, or whether extensive rewriting is necessary.

Development of self-correcting AMG and smoothed aggregation continues. Tests show that both methods can work well for a number of problems for which AMG is not currently well suited. At the moment, only fairly simple model problems have been studied, and the methods have not yet been implemented in FOSPACK.

Much has been accomplished in this six-month project, with the main advances the introduction of higher-order finite-element spaces and time-stepping capability. While more needs to be done here, this already allows much increased accuracy for invested work and the study of FOSLS methods applied to a wider range of problems.

Adaptive Numerical Methods for Reactive and Nonreactive Flow on Overlapping Grids

Donald W. Schwendeman
Rensselaer Polytechnic Institute

Summary

We develop an adaptive numerical method for the accurate calculation of high-speed reactive and nonreactive flows on overlapping grids. In the reactive case, the flow is modeled by the reactive Euler equations with an assumed equation of state and with various reaction rate models. The nonlinear hyperbolic partial differential equations are solved with an unsplit, shock-capturing scheme; a Godunov-type scheme computes fluxes and a Runge-Kutta error control scheme computes the source term modeling the chemical reactions. An adaptive-mesh-refinement (AMR) scheme has been implemented in order to locally improve grid resolution. The method may also be applied to nonreactive flow problems, in which case the source terms are simply set to zero. The code uses composite overlapping grids to handle complex flow geometries. It is part of the Overture-OverBlown framework of object-oriented codes, developed in collaboration with Bill Henshaw, David Brown, and other members of the Overture team within the Center for Advanced Scientific Computing (CASC). During FY 2003, we explored two-dimensional detonation-propagation problems using an ignition and growth reaction model with a mixture JWL equation of state, extended the method with AMR to handle three-dimensional reactive and nonreactive flows, and also extended the method to handle two-dimensional moving meshes.

An ignition and growth reaction model is one of several reaction models now available in the reactive Euler code. This reaction model is designed to handle heterogeneous explosives and assumes a two-component mixture consisting of a (solid) reactant and a (gaseous) product. Each component is assigned a JWL-type equation of state. The reaction rate for the ignition and growth model involves several steps, each with its

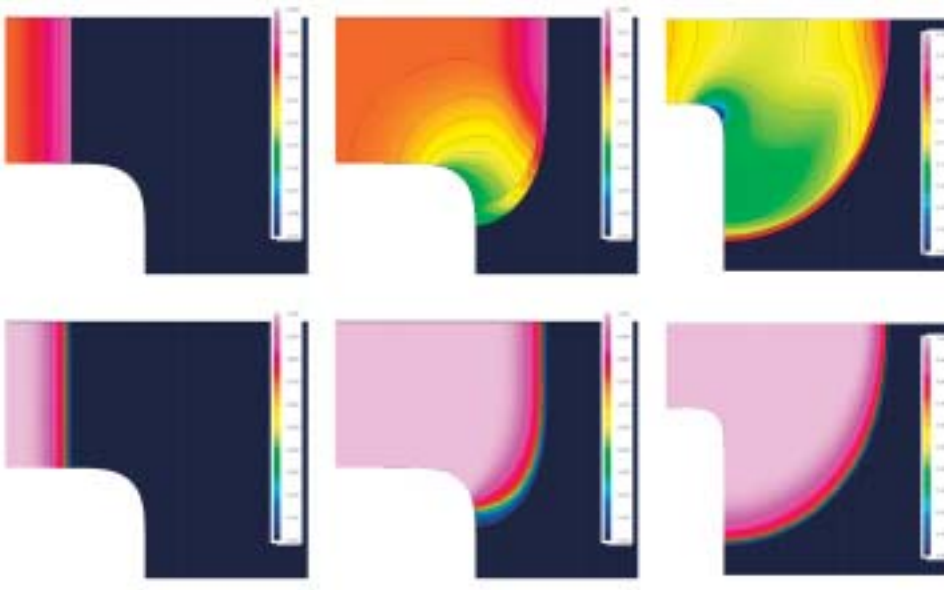


Fig. 1: Corner turning for an ignition and growth reaction model. Top row is pressure and bottom row is reaction progress.

Continued

Donald W. Schwendeman
Rensselaer Polytechnic Institute

Summary continued

own rate, typically pressure-dependent, and its own depletion law. It was used to study various problems involving detonation diffraction. One such problem involves detonation corner turning, as illustrated in Figure 1. In this problem a steady Chapman-Jouget detonation is propagating in a solid explosive (from left to right in the figure). The parameters used in the ignition and growth model are those for PBX9502. Of particular interest in this problem is the behavior of the detonation as a result of the interaction with the corner, and whether the detonation would be weakened to the point of failure. This calculation uses a composite overlapping grid with approximately 40,000 grid cells on the base level and two additional levels of AMR grids with a refinement factor of 4, so that the reaction zone is well resolved on the finest grid level. The corner is replaced by a 90° circular arc, whose radius is of the order of the length of the steady CJ reaction zone. Thus, the corner remains sharp on the length scale of the flow, but the corner singularity has been removed. As the detonation turns the bend its leading shock weakens and the reaction zone behind it lengthens. The strength of the shock, however, never falls below the level of compression needed to maintain the ignition step of the reaction model, and thus the reaction does not fail. In fact, the calculations show that the detonation ultimately strengthens once it has completed its interaction with the bend and is propagating down the straight vertical wall.

An extension to handle three-dimensional reactive and nonreactive flows is currently under testing for both accuracy and efficiency. One such test involves shock diffraction by a sphere in a nonreactive gas, as illustrated in Figure 2. In this calculation, a planar shock propagates from left to right in a channel with rectangular cross section and is diffracted by solid sphere centered at the origin. The base grid consists of a Cartesian grid for the channel and two orthographic grids to handle the boundary of the sphere. One AMR level with refinement factor equal to 4 is used in order to locally increase the grid resolution. The outline of the AMR grids may be seen along with contour plots on perpendicular cut planes for the time t = 0.6 in Figure 2. At this time, the planar shock has just impacted the sphere, and a high compression region and a reflected shock have developed in front of the sphere. At later times, there are too many AMR grids to be shown clearly in the figure so that only the contours of density are shown. The solution is well represented on the grid and the contours of density show the expected behavior of the diffracted

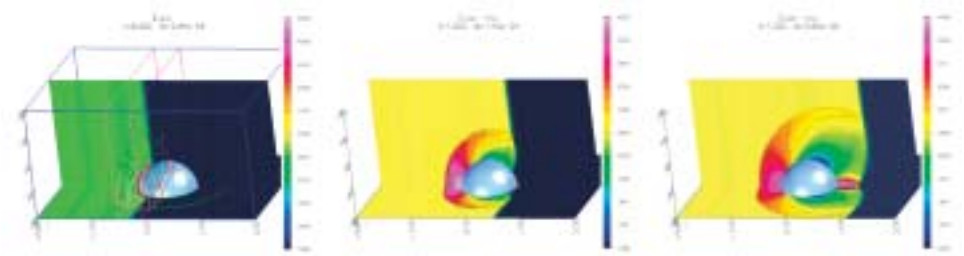


Fig. 2: Planar shock diffraction by a sphere. Density at t = 0.6, 1.0 and 1.3.

Continued

Donald W. Schwendeman
Rensselaer Polytechnic Institute

shock system. The contours of density also indicate that the numerical solution maintains axial symmetry even though this symmetry is not assumed in the fully three-dimensional calculation. This solution gives an indication of the problems being considered for testing. Other flows have been computed using the three-dimensional version of the code and have been tested for accuracy, but more testing is still needed, and some areas in which the code could be made more efficient have been identified, in particular the overhead needed to manage the AMR grids.

Work has begun to extend the code to handle discretizations involving moving meshes. The main issue here centers on the incorporation of a grid velocity into the numerical scheme. This involves modifications of both the calculation of the Godunov flux and the calculation of the source. The appropriate modifications to the flow-solver have been made and the work will now focus on testing the numerical method. One of our first tests is shown in Figure 3. In this test, a two-dimensional, nonreactive gas is at rest in a square channel and is disturbed at time *t=0* by the impulsive motion of an imbedded cylinder. The cylinder moves with a constant velocity from right to left. The motion of the cylinder is handled by the moving annular grid shown in the figure. The corresponding contours of density for the flow are shown above the grid. As the time-stepping proceeds, the grid overlap must be updated so that the grid generator is called for each time step. For sufficiently large grid velocities, the low-density region behind the cylinder presents a problem for the Roe approximate Riemann solver used in the calculation of the numerical fluxes. This problem is known in the literature and other approximate Riemann solvers are under development to fix it.

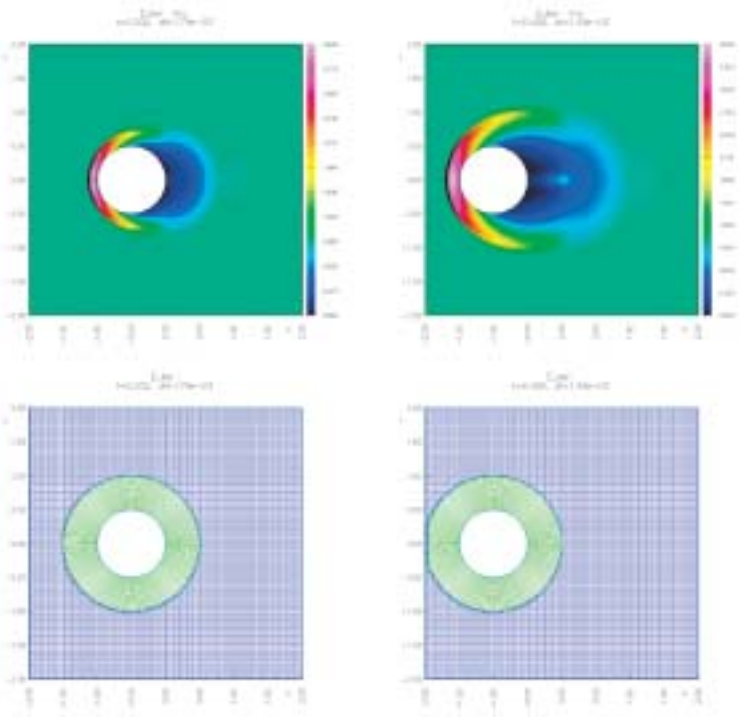


Fig. 3: Impulsive motion of a cylinder in a nonreactive flow.

Implicit Ocluders

Sinesio Pesco
Oregon Graduate Institutes
&
Claudio Silva
Oregon Graduate Institutes

ISCR Institute for Scientific Computing Research

Summary

Isosurfaces play a central role in the visualization of three-dimensional scalar fields. By being able to compute and display isosurfaces interactively, scientists can explore their datasets, study detailed features of interest, and obtain insights into the inner workings of real physical phenomena and simulated models. Because of their ubiquitous use in visualization, the computation and rendering of isosurfaces has received great attention from the visualization research community.

In this work, we propose a novel visibility-culling technique for optimizing the rendering and computation of opaque isosurfaces. Given a continuous scalar field $f(x)$ over a domain D and an isovalue w, our technique exploits the continuity of f to determine conservative visibility bounds implicitly, i.e., without the need for actually computing the isosurface. We generate implicit occluders based on the change in sign of $f^*(x) = f(x) - w$, from positive to negative (or vice versa) in the neighborhood of the isosurface. Consider, for example, the sign of f^* along a ray r cast from the current viewpoint. The first change in sign of f^* within D must contain an intersection of r with the isosurface. Any additional intersection of the isosurface with r is not visible. Implicit occluders constitute a general concept that can be exploited algorithmically in different ways depending on the framework adopted for visibility computations. In this work, we propose a simple "from-point" approach that exploits well-known hardware occlusion queries.

The implicit occluders project is in an advanced state of development. We first worked on a single resolution implementation that only worked for regular grids. Over the summer of 2003, we worked on extending this code to unstructured grids. While at LLNL, we have been working primarily on extending this work into an adaptive level-of-detail algorithm by extending the dual contouring work of Ju et al.

As of October 2003, I will be moving to the University of Utah. There, I will continue to work on this project. We would like to submit a paper on the implicit occluders work later this fall. Some options would be VisSym '04 or CGI (Computer Graphics International) '04. We are still deciding what material should be included in this paper, i.e., whether the adaptive work should be part of it, or simply send a revised version of the Visualization 2003 submission and leave the new adaptive work for a Visualization 2004 submission.

Institute for Scientific Computing Research

Workshops and Conferences

Workshops and Conferences

As a part of its outreach, the ISCR hosts and sponsors various workshops and conferences throughout the year. In many cases, the ISCR co-hosts or co-sponsors the event with other organizations internal and external to LLNL. Subject matter may be drawn from any of the wide range of research interests identified by the ISCR.

Workshop on Solution Methods for Large-Scale Nonlinear Problems

Synopsis of Workshop

Research advances in computational science technology have truly made simulation an equal partner with theory and experiment in understanding physical phenomena. Simulation is increasingly relied upon in government and industry for high impact decisions. As a result, computational scientists are called upon to solve problems with more complex and coupled physics than ever before. Applications such as fluid dynamics, fusion, electronics, groundwater flow, astrophysics, and combustion are producing coupled models where individual effects were previously studied. In addition, parallel computers with large-storage capacities have paved the way for high-resolution simulations on large-scale domains. The growth in complexity and size of models coupled with the advent of more powerful machines has led to a continued and increasing demand for effective algorithms for solving large-scale systems of nonlinear equations.

The Workshop on Solution Methods for Large-Scale Nonlinear Problems, held August 6–8, 2003, in Livermore, California, brought together many of the most active researchers on both the algorithmic and the applications sides of the area. The workshop was hosted by the Center for Applied Scientific Computing and the Institute for Scientific Computing Research at Lawrence Livermore National Laboratory. Major themes included Newton–Krylov methods, preconditioning techniques, operator-split and fully implicit schemes, continuation methods, and Partial Differential Equation (PDE)-constrained optimization. Applications addressed included fluid dynamics, astrophysics, magnetohydrodynamics, porous media flows, radiation diffusion, and other areas. The workshop was a follow-on to similar workshops on large-scale nonlinear problems held at Utah State University (1989, 1995) and in Pleasanton, California (2000). In the spirit of the earlier workshops, liberal break times in the schedule of talks and discussion sessions encouraged informal interactions among participants. Workshop attendees, mainly from the United States, came from academia (17), government laboratories (25), and industry (2). Each of the three workshop days consisted of several talks followed by a moderated discussion of selected issues.

From the presentations, it was clear that the workhorse methods in the field are still Newton-Krylov methods, in which Newton's method is combined with preconditioned Krylov methods for approximately solving the linear Jacobian systems. Efficient schemes for achieving fast convergence, up to the quadratic convergence of Newton's method, have been realized in a number of applications, leaving "robustification" as the main focus of current research. In particular, a number of workshop talks concentrated on globalization and continuation techniques combined with Newton-Krylov methods. The most commonly applied globalization has been *linesearch* (*backtracking*, *damping*), in which each step direction is that of the approximate solution of the Jacobian system and the step length is chosen to give desirable progress toward a solution. *Trust region*

Continued

Workshop on Solution Methods for Large-Scale Nonlinear Problems

Synopsis of Workshop continued

methods offer an alternative approach, in which each step is taken within a "region of trust" of the local linear model around the current approximate solution. Trust-region implementations of Newton's method combined with the conjugate gradient method (known as *truncated Newton methods*) have been applied to large-scale optimization problems for some time. In contrast, trust-region globalizations of more general Newton-Krylov methods have not often been used in the past; however, they have recently received attention. Pseudo-transient continuation methods, which have been used in computational fluid dynamics for some years, are now starting to be invoked in other application areas such as groundwater flow. Similarly, parameter continuation methods, which have been used for some time in certain applications, are enjoying broader usage in applications with Newton's method. Systematic comparison of all of these robustification techniques has begun, and further understanding of which method is best for a particular problem is a subject of active research.

Other workshop talks discussed new advances in the use of Newton's method. One such advance is the development of theory for extensions of the method to certain classes of non-smooth nonlinear functions. Using the concept of generalized derivatives, convergence of Newton's method has been shown for these functions, both with and without pseudo-transient continuation. Another advance is in the use of automatic differentiation for efficiently generating accurate Jacobian-vector products within a Newton-Krylov method; this has shown computational speed advantages over finite-difference approximations for fluid dynamics problems. Additionally, initial results were given for applying a two-grid technique that transfers nonlinearities to a coarser resolution version of the original problem. Lastly, there have been new developments in globalized tensor-Krylov methods, which can be regarded as extensions of Newton-Krylov methods that incorporate limited second-order information. Recently developed globalizations, in combination with certain extensions of the Generalized Minimal Residual (GMRES) method, provide robustness and efficiency while allowing the superlinear convergence associated with direct tensor methods on singular and ill-conditioned problems.

Preconditioning for both nonlinear and linear problems continues to be an important area of research. Further advances in preconditioning the nonlinear problem via a nonlinear additive Schwarz approach were outlined, including three-dimensional and parallel results on computational fluid dynamics problems. Preconditioning the linear Jacobian systems continues to be a primary requirement for the success of Newton-Krylov methods on large-scale problems. An application that had significant interest at the workshop and which has benefited from advances in preconditioning is fusion simulation. Application of Newton-Krylov methods along with the development of effective preconditioning has allowed a move to implicit formulations for some problem classes in this field. These formulations allow for larger time steps that are no longer

Continued

Workshop on Solution Methods for Large-Scale Nonlinear Problems

Synopsis of Workshop continued

limited to tracking the fastest wave in the system. Extensions of these approaches were also given for the more complicated case of implicit formulations of fusion problems on adaptive meshes. In astrophysics, sparse approximate inverse preconditioners have been developed for multigroup flux-limited diffusion of neutrinos for core-collapse supernovae simulations. These preconditioners have enabled use of Newton-Krylov methods and thus larger time steps for these problems.

Talks from applications scientists reflected the incorporation of these algorithmic advances into scientific simulation codes. In particular, groundwater and geomechanics codes are being coupled together as are shallow water and groundwater simulators. Although these couplings are not all yet fully implicit, nonlinear couplings between the relevant effects can now be studied. In the case of fully implicit formulations of shallow water problems, improved preconditioning and nonlinear solvers allow simulation of hurricanes with much larger time steps than previously could be used. In other applications, preconditioned Newton-Krylov methods have enabled progress in simulation of combustion chemistry and also in modeling phase transitions in smart materials that previously were intractable.

Advances in Newton-Krylov methods have also benefited shape optimization applications, in which the typical PDE solver is now part of a constraint that must be completed for each iteration of the optimization method. Advances in this field have developed by viewing the optimization problem as a large system where the optimization and constraints are handled “all-at-once” with a Newton-Krylov method applied to a system comprised of the objective and the PDE constraints.

Each of the three workshop days ended with a moderated discussion. The discussion on the first day dealt with robustness and failure of nonlinear solvers. The session began with a review of types of failure, including divergence or stagnation of the iterates, convergence to a local norm minimizer that is not a solution, failure of the linear solver to produce adequate linear residual reduction, and convergence to a “wrong” solution, such as a steady state that is dynamically unstable or which violates physical constraints. This was followed by a discussion of causes, symptoms, and possible remedies for various failure modes, together with the general question of how to construct nonlinear solver algorithms to prevent failure if possible and to terminate with useful diagnostic information otherwise. The session concluded with a brief discussion of experiences with several globalization methods (including continuation) and their relative effectiveness in practice.

The second day's discussion was on tolerances, stopping criteria, and related accuracy/efficiency considerations for nonlinear solvers. This discussion initially focused on fully implicit methods for time-dependent problems, first reviewing their merits and

Continued

Workshop on Solution Methods for Large-Scale Nonlinear Problems

Synopsis of Workshop continued

disadvantages relative to explicit and semi-implicit methods and then addressing the issue of stopping tolerances for the nonlinear residuals associated with the implicit equations that are sufficient to ensure desired accuracy. In general, participants felt that a relative stopping criterion such as 10^{-5} on the nonlinear residual norm worked well in most instances, although some preferred absolute tolerances such as 10^{-5} or 10^{-6} in some applications. The next focus of the discussion was on choosing the difference step in finite-difference approximations of Jacobian-vector products in "matrix-free" Newton-Krylov implementations. Participants commented on their experiences with several previously proposed choices, with some agreement that difference-step formulas should take into account inner products as well as norms of the relevant vectors. In conclusion, the discussion touched on the issue of choosing the relative residual norm tolerances (the *forcing terms*) for approximately solving Jacobian systems in order to enhance efficiency and perhaps robustness as well. Participants reported a variety of experiences, with effective choices in particular applications ranging from adaptive formulations to relatively large (10^{-1}) and small (10^{-4}) constant choices.

The third day's discussion was on algorithms and architectures. The discussion centered on what nonlinear solver software writers and algorithm designers need to pay attention to in order to take advantage of upcoming architectures. The two main consensus items for this question were: careful use of data, so that cache hierarchies can be exploited and data reuse maximized; and fault tolerance in implementations. It was noted that with the large numbers of processors on platforms such as the new ASCI machines and Blue Gene/Light, it will be likely that some number of processors will fail and drop out of a simulation unexpectedly. Implementations of algorithms for these machines will need to tolerate these faults and robustly solve for solutions.

A poster session held on the evening of the second day highlighted the work of students attending the meeting as well as software packages offering robust implementations of Newton-Krylov methods and other nonlinear solvers. The student posters reflected a breadth of work in the field, including solvers for electrical tunneling, optimization methods for groundwater remediation, globalization of Newton-Krylov methods for the Navier-Stokes equations, and variants of Newton-Krylov methods for problems with expensive nonlinear function evaluations. The software posters provided information on the Sandia package NOX, the Lawrence Livermore package SUNDIALS, and the suite of packages offered through the TOPS (Terascale Optimal PDE Simulations) SciDAC project including the Argonne package PETSc. All of these packages are available to the public and offer various functionalities in addition to Newton-Krylov solvers.

Although the presentations at the workshop reflected significant advances in the field of nonlinear solvers, many challenges still remain. Robust solvers for problems discretized with discontinuous Galerkin methods continue to be an open research area, as does

Continued

Workshop on Solution Methods for Large-Scale Nonlinear Problems

Synopsis of Workshop continued

productive use of multiple-grid levels in nonlinear solvers. Development of nonlinear multigrid and two-grid methods has shown promise, but the robustness and efficiency gains seen on linear problems still have not appeared in the nonlinear realm. Fault tolerant implementation of nonlinear solvers also remains as an important open area.

The field of solution methods for large-scale nonlinear problems is an active research area vital to simulation technology. Advances in nonlinear solver algorithms are helping to push back the frontiers of science in many applications. Interested researchers can see more of this field in sessions on nonlinear solvers at the Copper Mountain Conference on Iterative Methods March 28–April 2, 2003, and at a mini-symposium (submitted) at the Society of Industrial and Applied Mathematics (SIAM) 2004 annual meeting. More information about the workshop described above can be found at www.llnl.gov/casc/workshops/nonlinear_2003 where abstracts and PDF files for many presentations are archived.

Advanced Scientific Computing Requirements Workshop

Synopsis of Workshop

The Advanced Scientific Computing Requirements Workshop was held October 8–9, 2003, at the Crystal City Hilton in Arlington, Virginia. Approximately 95 participants, representing the U.S. Department of Homeland Security (DHS), national laboratories, federal agencies, academia, and industry convened to learn about DHS mission needs and identify mathematics and computer science research and development topic areas to address them.

The Advanced Scientific Computing (ASC) R&D program is part of the Science and Technology (S&T) Directorate within DHS. This program funds R&D to develop enabling computational technologies that will be deployed in the next generation of operational tools designed to simulate and understand physical and biological phenomena for homeland security applications. The ASC program also funds R&D for large-scale data management and integration. These capabilities will be used to improve scenario planning and emergency response, to design better radiation and biological detectors for border protection, to enable vulnerability assessments and consequence analyses for infrastructure protection, and to accelerate information extraction and data assimilation for threat analysis and attribution.

The workshop was organized around six "portfolios" in the S&T Directorate that exist to lead technology development for the operational directorates in DHS. These are the Borders and Transportation Security (BTS) Portfolio, the Emergency Preparedness and Response (EP&R) Portfolio, the Critical Infrastructure Protection (CIP) Portfolio, the Threat Verification, Testing, and Assessment (TVTA) Portfolio, the Chemical and Biological Countermeasures (ChemBio) Portfolio, and the Radiological and Nuclear Countermeasures (Rad/Nuc) Portfolio. Each day of the two-day workshop was structured to provide presentations associated with three of the portfolios in the morning followed by breakout sessions in the afternoon to consider R&D requirements for those portfolios.

Each morning's presentations included overviews for the day's portfolios, talks by DHS operational directorate staff to illustrate mission activities that rely on advanced technology, and presentations by R&D technologists working in applicable technology domains. The afternoon breakouts were organized along portfolio lines but further subdivided to emphasize either modeling and simulation or data sciences. In the breakouts, DHS technology users engaged in roundtable brainstorming with technology developers to identify and refine technology requirements for DHS operational missions.

Going into the workshop planning stage, the ASC program had already identified DHS needs for modeling and simulation and various data-science-related technologies. The workshop served to materially refine and expand these needs and enabled the development of a better sense of the impact that various technologies will have across

Continued

Advanced Scientific Computing Requirements Workshop

the mission space of DHS. Several data-science technologies—including information sharing and dissemination, data fusion and heterogeneous database integration, and information extraction—were revealed to have across-the-board utility within the portfolios. Likewise, discrete event simulation technology and visualization/visual analytics were needs shared among all of the portfolios. Other technologies had widespread if not universal applicability within the portfolios, including continuum simulations, data-driven simulations, decision support technology, collaborative technology, and geographic information systems. All results from the workshop are being compiled into a final report that will present conclusions as well as more detailed discussions of research opportunities for the ASC program.

Workshop participants, DHS representatives in particular, considered the event a great success, so much so that the workshop is looked on as a model for future requirements development activities. This outcome follows directly from the thorough preparatory efforts in the weeks preceding the workshop by the Program Committee, the Organizing Committee, and the Krell Institute, which was contracted to manage logistics and support planning activities.

As the ten-member Program Committee (senior scientists and technology managers from federal laboratories, academia, and industry) developed the workshop agenda and identified respected speaker and participant pools, the Organizing Committee (LLNL and Krell Institute staff) completed an intensive schedule of trips to Washington DC to interview portfolio managers and their staff. The tangible results of these interviews were the portfolio overviews prepared by the Organizing Committee and presented by portfolio managers at the workshop. Less tangible, but no less important, was the knowledge-base developed of the mission needs and potentially applicable technology domains that permitted the workshop to begin with the primary focus areas already identified. This conferred considerable benefit by maximizing workshop productivity, an important consideration given that the duration of the workshop was necessarily limited to minimize the time commitment from participants.

The Krell Institute, in addition to participating in developing the content of the workshop, handled all logistics for the workshop and also developed the workshop Web site. Krell staff were highly responsive in all tasks and generously lent their expertise wherever they could add value, before, during, and after the workshop.